

Received 19 January 2018.

Accepted 25 April 2018.

AN EXPERIMENTAL INVESTIGATION OF VARIATION IN SPANISH DIMINUTIVES

David Ellingson EDDINGTON

Brigham Young University *

eddington@byu.edu

Abstract

In an online survey, 656 Spanish speakers from 21 countries were asked to select between the long and short diminutive forms of 100 base words (e.g. *novio* > *novito*, *noviecito*). The influence of the base word, the base word's frequency, and the country of origin, gender, age, and educational level of the speakers were assessed. The most salient finding is that words have different diminutive forms depending on the country the speaker is from. However, a great deal of variation exists within each country, and few participants prefer the same diminutive form for all base words with a similar structure. Moreover, the influence of age, gender, and educational level is also apparent for certain classes of words. Country-specific analyses of the data from Spain and Mexico were also performed which confirms the influence of social factors on diminutivization.

Keywords

Spanish, diminutive, experimental approach

* Brigham Young University 4064 JFSB, Provo, Utah 84602, USA.

UNA INVESTIGACIÓN EXPERIMENTAL DE LA VARIACIÓN EN DIMINUTIVOS ESPAÑOLES

Resumen

En una encuesta en línea, se pidió a 656 hispanohablantes de 21 países que seleccionaran entre las formas diminutas largas y cortas de 100 palabras base (por ejemplo, *novio* > *novito*, *noviecito*). Se evaluaron la influencia de la palabra base, la frecuencia de la palabra base y el país de origen, género, edad y nivel educativo de los hablantes. El hallazgo más destacado es que las palabras tienen diferentes formas diminutas según el país de origen del hablante. Sin embargo, existe una gran variación dentro de cada país, y pocos participantes prefieren la misma forma diminuta para todas las palabras base con una estructura similar. Además, la influencia de la edad, el género y el nivel educativo también es evidente para ciertas clases de palabras. También se realizaron análisis específicos de los datos de España y México que confirman la influencia de los factores sociales en la diminutivización.

Palabras clave

español, diminutivo, aproximación approach

1. Introduction

In Spanish, diminutives are used to indicate concepts such as small size, familiarity, lack of importance, cuteness, and disdain (Ruiz de Mendoza 1996, Zuluaga 1993). A number of diminutive suffixes exist (e.g. *-uelo*, *-ico*, *-eta*, *-illo*), but the most frequent diminutive suffix in Spanish is *-ito*, which is why it is the focus of the present study. The short diminutive suffix *-ito* and its feminine counterpart *-ita* have long allomorphs as well (e.g. *-cito/a* and *-(c)ecito/a*). The distribution of these allomorphs has been the subject of several theoretical analyses (Ambadiang 1996, 1997; Bermúdez-Otero 2007, 2013; Bradley & Smith 2011; Castillo & Ortiz 2013; Castro 1998; Colina 2003; Crowhurst 1992; Eddington 2002; Elordieta & Carreiras 1996; Horcajada 1988; Miranda 1999; Norrmann-Vigil 2012; Prieto 1992; Reynoso 2005; Rojas 1977; Smith 2011; Stephenson 2004). These studies describe diminutivization within different frameworks, the details of which are not relevant to focus of the present study, and a review of which would occupy a lengthy paper on its own. A number of other studies focus on diminutives in one particular variety (Bradley & Smith 2011; Crowhurst 1992; Fontanella 1962; Gaardner 1966; Jaeggli 1980).

However, in spite of all the attention diminutivization has received, cross-dialectal variation in the formation of diminutives has been the topic of only a few studies. Among these, Prieto (1992) gathered intuitions from one or two speakers from seven countries, while Callebaut (2011) extracted diminutives from 14 countries using the CREA corpus¹ and other online sources. Clearly, more pan-Hispanic data is called for.

One of the principal aims of the present study is to gather more comprehensive data that includes diminutives from all Spanish-speaking countries. The study was designed to be deep rather than broad. That is, instead of obtaining a few speaker intuitions about a large number of words in a given country, input from many speakers on a smaller number of words was the aim. These data were used to determine how much variation exists in a single country. That is, are the diminutives of base words with similar phonological structures formed in the same way in a given country? Which countries use more short or long forms of diminutives? In like manner, it will be possible to determine if individual speakers form diminutives with similar bases in the same way, or if on the contrary, diminutivization is carried out in a word-by-word fashion.

The role of word frequency will also be tested since the test items included both highly frequent and highly infrequent bases. Another void in the diminutive literature, which the present paper will address, is how social factors such as age, gender, and education may influence diminutive formation. Besides determining the effect these factors have on diminutives across the Spanish-speaking world, the study will also show how the resulting data may be used to investigate diminutive processes in a particular country. To this end, diminutive formation is explored in Mexico and Spain, two countries that had a large number of survey participants.

2. Method

Participants were first solicited via posts on Facebook as well as on Mechanical Turk. This resulted in 710 responses, of which the data from 54 participants were

¹ corpus.rae.es/creanet.html.

eliminated, either because the participants answered the truth tester words incorrectly, or indicated that Spanish was not their native tongue, or that they were under 18 years of age. This left 656 survey on which to base the analyses. Biographical data on the participants was gathered: age, gender, country of origin, level of education (primary school, secondary school, college), and whether or not Spanish was their native language. The instructions for the survey read:

En español es bastante común usar palabras diminutivas como 'agüita' en vez de 'agua' o 'rapidita' en vez de 'rápida.' Sin embargo, hay diminutivos que difieren de país en país. Por ejemplo, para algunas personas el diminutivo de 'café' es 'cafecito' y para otras es 'cafelito' o 'cafetito.' Los tres se consideran correctos dentro del país en que se usan.

A continuación vas a ver una lista de palabras con dos posibles diminutivos. Escoge el diminutivo que te suene mejor en tu propio país. Habrá palabras poco comunes que puedes desconocer. Escoge un diminutivo aunque no conozcas la palabra.

At this point they were presented survey items in randomized order in the following format, and asked to choose between the two responses:

¿Cuál es el diminutivo de timbre?

☐ timbrecito

☐ timbrito

The response order was randomized for each participant as well. All of the participants were asked to respond to the 28 common words, and to the three truth tester words, that are marked with asterisks in Appendix 1. Fifty-two test items were also randomly chosen from the remaining test items and presented to the participants. Words from these two groups were combined and presented in a different random order to each participant. Each participant gave their input on a total of 83 test items. Restricting the

survey to 83 test questions the only way to design a test that required only 10-15 minutes to complete.

2.1 Participants

The number of participants varied widely between countries (Table 1). Uneven data of this sort is common in sociolinguistic studies of this nature. Of the total 656 participants, 368 were male and 288 female, with 271 being between 18 and 29 years of age, 193 between 30 and 39, 94 between 40 and 49, and 98, 50 and older. Thirteen participants indicated having some primary education or having finished primary school, 172 finished secondary school, and 471 had a college degree.

Country	AR	BO	CL	CO	CR	CU	DO	EC	ES	GT	HN
#	29	8	18	43	10	10	24	12	168	11	6
Country	MX	NI	PA	PE	PR	PY	SV	US	UY	VE	
#	96	3	3	19	13	1	5	23	16	138	

AR Argentina, BO Bolivia, CL Chile, CO Colombia, CR Costa Rica, CU Cuba, DO Dominican Republic, EC Ecuador, ES Spain, GT Guatemala, HN Honduras, MX Mexico, NI Nicaragua, PA Panama, PE Peru, PR Puerto Rico, PY Paraguay, SV El Salvador, US United States, UY Uruguay, VE Venezuela

Table 1. Number of participants from each country.

2.2 Test items

In the majority of Spanish words that undergo diminutivization with the suffix *-ito/a*, there is no variation. For instance, the diminutive of *mesa* 'table' is *mesita* and the diminutive of *zapato* 'shoe' is *zapatito* universally. However, a great deal of variation occurs in certain kinds of words. The present paper focuses on four word classes that demonstrate variation (Real Academia Española 2009: 643-651):

(1) monosyllabic words (e.g. *tren* 'train' > *trencito/trenecito*, *rey* 'king' > *reycito/reyecito*)

- (2) bisyllabic words ending in *-o* and *-a* with the diphthongs *-ie-* and *-ue-* in the stem (e.g. *viejo* ‘old’ > *viejito/viejecito*, *pueblo* ‘town’ > *pueblecito/pueblito*)
- (3) bisyllabic words ending in *-e* (e.g. *dulce* ‘sweet’ > *dulcecito/dulcito*, *diente* ‘tooth’ > *dientecito/dientito*)
- (4) bisyllabic words ending in *-io/a* (e.g. *indio* ‘Indian’ > *indiecito/indito*, *rubio* ‘blonde’ > *rubiecito/rubito*).

A total of 100 test items were used in the study (see Appendix 1): 35 *-io/a* final words (e.g. *bestia*, *necio*), 29 words with *-ie-*, *-ue-* in the stem which end in *-o* or *-a* (e.g. *duelo*, *tienda*), 29 *-e* final words (e.g. *torpe*, *carne*), and 4 monosyllabic words (e.g. *pie*, *tren*). Words in these categories were chosen that ranged from highly frequent to infrequent based on the Clearpond Corpus.² In addition to these test words, three other truth tester words were included as a way of determining if a participant was truly paying attention to the test items in the survey or not. When someone chose the non-existent diminutives in these truth tester word (i.e. *manzanecita*, *cancionito*, or *arbolita*) the validity of their responses was questioned and all of that participant’s responses were summarily eliminated from the analyses. For each test item, two responses were provided for participants to choose from, a short and a long form (e.g. *bestia* > *bestita*, *bestiecita*; *duelo* > *duelito*, *duelecito*; *torpe* > *torpito*, *torpecito*; *pie* > *piecito*, *piececito*).

2.3 Test variables

The influence of a number of variables was tested including age, education, gender, and test item. In order to elicit their country of origin the participants were asked: *¿De dónde eres? (O ¿a qué país pertenece tu manera de hablar?)* In addition, some linguistic variables were considered as well: the log frequency of the base word, the number of orthographic neighbors it has, and its CV structure. However, none of these linguistic

² clearpond.northwestern.edu/spanishpond.html.

variables reached significance in any analysis, so they are not discussed further. All analyses were mixed effects logistic regressions with participant as a random effect.³

Test words were chosen based on their frequency in order to test a particular hypothesis. Highly frequent diminutives that have been produced or heard many times in the past are more likely candidates for whole word storage in the mental lexicon rather than receiving their diminutive via some online process of diminutivization. Low frequency words, on the other hand, especially extremely unusual words that speakers are unlikely to have diminutivized before, are much more likely to be derived by some online process. Both high and low frequency words were included for this reason. If differences are found between high and low frequency items, that may suggest different processing strategies. The default diminutivization process would be assumed to apply to the low frequency forms.

3. Results for all countries combined

3.1 Results for monosyllabic words

The four monosyllabic words that were included in the survey (i.e. *tren*, *rey*, *pie*, *té*) have long forms ending in *-(c)ecito*,⁴ and short forms ending in *-cito*. The only significant variables that influenced the use of each allomorph were age⁵ ($F(1, 2525) = 22.178$, $p < .0005$) and the interaction between test word and country ($F(60, 2525) = 3.84$, $p < .0005$). The positive coefficient of .299 for the age variable indicates that as participants become older, they use more long forms such as *piececito* and *trenecito*, while younger speakers favor *piecito* and *trencito*.

³ Including a random effect for words either caused errors in the Hessian matrix, or produced non-convergence.

⁴ *-cecito* appears after vowel final bases, and *-ecito* after consonant final bases.

⁵ Because age was elicited in the survey by decade, the variable was originally categorical. This created a separation in the data since participants from each group were not found in every country. This was overcome by converting the categorical variables into continuous variables in this analysis.

The relationship between the test item and country appears in Table 2, which summarizes the results of an LSD post hoc analysis that is based on the overall mean of the words for comparison. This mean is used to calculate whether the country specific mean for the word differs significantly from the overall mean ($p < .05$). An 'S' indicates that the short form was preferred significantly more than the overall mean, and an 'L' shows that the long form was preferred significantly more than the overall mean. Lower case letters are given to countries that only had only 1-9 participants, in order to show that results that are based on fewer participants may less accurately reflect diminutive formation in a particular country. Blank cells indicate no significant preference for either long or short forms. The last two rows show the sum of the number of significant long and short forms in each country. It is a rough estimate of which countries prefer each form. It is clearly evident that the long forms are a characteristic of Peninsular Spanish and to a lesser extent Mexico. The sole participant from Paraguay, on the other hand provided short forms for all four of these test words.

Word	AR	BO	CL	CO	CR	CU	DO	EC	ES	GT	HN	MX	NI	PA	PE	PR	PY	SA	US	UY	VE
<i>pie</i>	S	s	L						L	S	s						s	s		S	
<i>rey</i>								S	L	L		L	s	s			s				
<i>té</i>	S	s	S	S	S		L	S	L		s	S		s	S	L	s	s		S	S
<i>tren</i>	S	s	S	S	S		S	S	L		s	L	s	s	S		s		L	S	S
# of long	0	0	1	0	0	0	1	0	4	1	0	2	0	0	0	1	0	0	1	0	0
# of short	3	3	2	2	2	0	1	3	0	1	3	1	2	3	2	0	4	2	0	3	2

Results based on 2-9 participants are in lower case.

Table 2. Relationship between country and long and short versions of monosyllabic test items.

Although it is the interaction between test word and country which reached significance in the analysis, it is useful to observe the raw proportions. Figure 1 contains the proportions of long and short diminutives by country, and illustrates how the long forms are much more prevalent in Spain (Real Academia Española 2009: 648), followed at a distance by Mexico. In contrast, the countries on the bottom of the chart favor the short forms of these monosyllabic words.

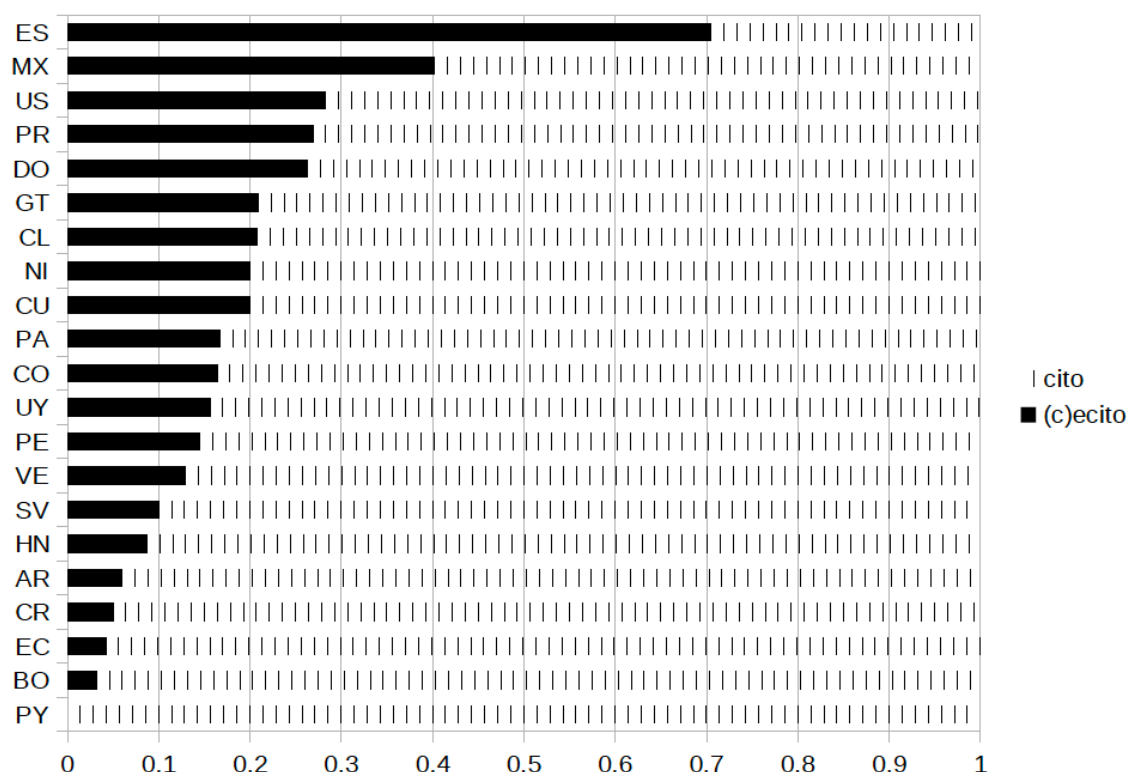


Figure 1. Proportion of long and short responses to monosyllabic words by country.

We know that much of the variation observed in a single country is due to different speakers producing different diminutives. One question that remains is the role that single speakers play in the variation. It may be that some speakers produce either all short forms for monosyllabic diminutives, or all long forms. If speakers produce a mixture of long and short forms, then some of the variation in each country is due to individual speakers as well. What the data show is that only 34% of the participants consistently chose all long forms while 5% chose all short forms. The remainder preferred some mixture of the two, which means that much of the variation within a country is attributable to the individual speakers rather than to variation across different speakers alone.

3.2 Results for words with diphthongs

This category of test words comprises 29 words containing the diphthongs *-ie-*, *-ue-* in the stem, which also end with either *-o* or *-a*. Only gender ($F(1, 15998) = 7.02, p < .008$) and the interaction of test word and country ($F(555, 15998) = 2.148, p < .0005$) were found to influence the choice of diminutive allomorph.⁶ The coefficient for gender ($-.302$) shows that women slightly disfavored the short forms (i.e. preferred the long diminutives) when compare to men (Short form: 53.2% women versus 55.1% men.) Women often reflect more prestigious forms of speech, and these data lead one to wonder if longer forms are either tacitly more prestigious or possibly are the ones taught in school. Clearly more study is needed in this regard. However, it could be argued that Peninsular Spanish is the source of prestige since the following section describes the long forms as most typical of that variety.

Eddington (2002) hypothesized that all words of a particular type (e.g. containing *-ie-*, *-ue-*) would be diminutivized in the same way in a given dialect. The country by test item interaction that appears in Table 3 refutes this hypothesis. In that table, blank cells indicate that there is no statistical preference for either a long or short word in a particular country, meaning that both forms are used to some degree. ‘S’ indicates that the short form was preferred significantly more than the overall mean, and an ‘L’ shows that the long form was preferred significantly more than the overall mean. Therefore, speakers in most countries appear to use some mix of short and long forms which supports the idea that diminutivization is word and country specific.

As far as extremes are concerned, the counts at the bottom of the table suggest that Peninsular Spanish once again stands out from other varieties in its strong preference for long forms such as *piedrecita* and *quietecito* over *piedrita* and *quietito* (Real Academia Española 2009: 643). On the other end of the spectrum, Argentina and Uruguay, which comprise the Rioplatense dialect, prefer the short diminutives for the majority of the test words. This same finding is observed in the proportions of long and

⁶ One variable that was insignificant was whether the word contained *-ie-* versus *-ue-*.

short diminutives in each country (Figure 2). Spain appears at the top and Argentina and Uruguay at the bottom.

Word	AR	BO	CL	CO	CR	CU	DO	EC	ES	GT	HN	MX	NI	PA	PE	PR	PY	SA	US	UY	VE
<i>ciervo</i>	S						S	S	L				s	s			s		L	S	
<i>cuerpo</i>	S	s							L	L		L					l	l		S	
<i>diestro</i>	S								L					l			l	l			L
<i>diezmo</i>					S			S					l				l				
<i>duelo</i>	S	s							L				s				s	s		S	
<i>fieltro</i>				L		L			L		l			l		S	s				
<i>fiera</i>	S								L				l			L		s			
<i>fiesta</i>	S	s		L		L	L	S	L	L	l	L	l	l			s	l		S	L
<i>huevo</i>	S		S	S	S		S	S	L		s		s	l			s	s		S	
<i>juego</i>	S	s			S				L	S	s			s	S		s	s		S	S
<i>lienzo</i>	S		S	L	S				L			L	l	s			s				
<i>mielga</i>									L		s			l			l	s		S	
<i>mierla</i>			L						L			L	s		S		s			S	
<i>mueca</i>		s			S	S			L		s		s				l			S	S
<i>nieto</i>	S	s		L			L		L	L	l	L	l	l		L	s	l		S	L
<i>nuevo</i>	S	s		L		L	L	S	L		l	L	l	l		L	l		L	S	L
<i>piedra</i>	S	s				L	L		L			S					s	s		S	S
<i>pierna</i>	S	s	L	L	S				L	S	s	S	s				s	s			S
<i>puesto</i>	S	s	S	L		L	L	S	L				l		S		s	l		S	L
<i>quieto</i>	S						L	S	L		l	L		l		L	s			S	L
<i>riego</i>					L				L		l		l								S
<i>rueda</i>	S	s			S		S		L					s	S		s	s			
<i>ruego</i>	S			S			S		L	L			l				l	l		S	S
<i>siembra</i>							S	S	L					l			l				S
<i>suero</i>	S				S		S	S	L				s				s			S	S
<i>tienda</i>	S	s	S	S		L		S	L		s	S	s	s		S	l	s		S	S
<i>tuerto</i>	S		S			L			L		s		l							S	
<i>viejo</i>	S		S		S	S	S		L			S	s	s			l		S		S
<i>vuelta</i>	S	s						S	L		l	L			S		s	l	L	S	L
# of long	0	0	2	7	1	7	6	0	28	4	7	8	10	9	0	4	10	6	3	0	7
# of short	22	13	6	3	9	2	7	11	0	2	7	4	9	6	5	2	16	6	1	19	10

Results based on 1-9 participants are in lower case.

Table 3. Relationship between country and test items with diphthongs.

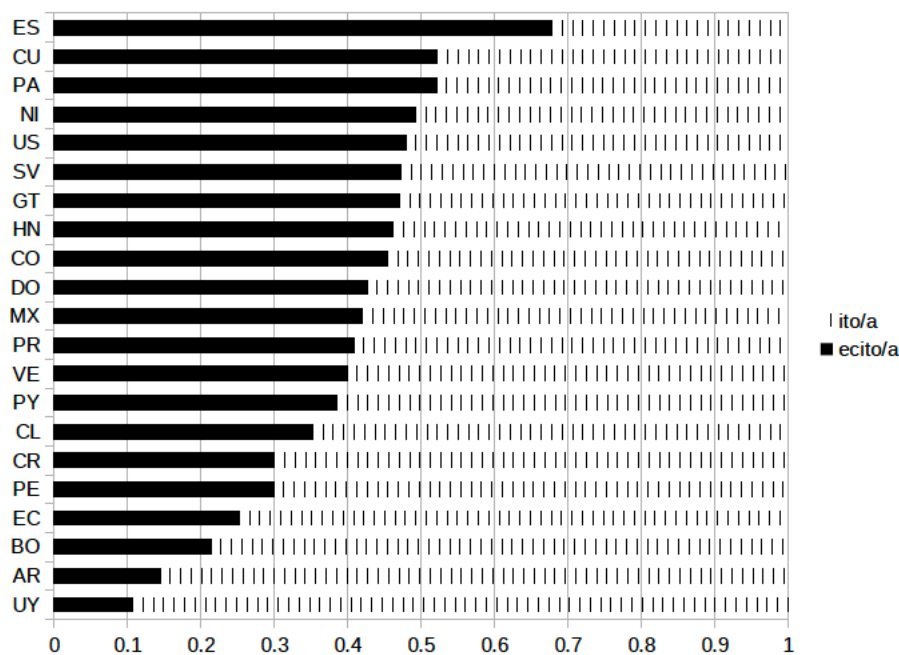


Figure 2. Proportion of long and short diminutive forms of words with diphthongs by country.

Besides giving insight into diminutive formation by country, the data allow us to consider individual variation as well. For example, to what extent does a single speaker prefer the long or short form of all diminutives of this sort? The answer is that only 18 participants preferred only short forms, while 20 preferred the long forms of all diphthongs containing *-ie-*, *-ue-*. The remaining 618 participants produced both long and short forms.

3.3 Results for words ending in *-io/a*

The 35 test words in this category include *rubio* and *bestia*. The long forms of these words were most prevalent comprising 82.3% of the cases. Only country by test item was significant in the analysis ($F(612, 16929) = 1.69, p < .0005$).⁷ Countries that strongly prefer mostly long or short diminutive forms are not clearly apparent in Table 4, while Figure 3 is somewhat more helpful. Panama and Costa Rica top the chart for the most long forms

⁷ There was separation in the data (knockouts) caused by a number of cells containing zeros in the data from the one Paraguayan and three Nicaraguan participants. This was solved by deleting the 27 instances from the Paraguayan participants and the 59 responses from the Nicaraguans prior to running the analysis.

while Ecuador and the US fall on the bottom. The data in Table 4 also serves to highlight diminutive forms in one country that are unusual in most others. For example, the diminutive of *voltio* in Cuba, for example, is more often *volitecito* than *voltito*. In the remainder of the countries it is either a statistical toss up between *volitecito* and *voltito* or the short form *voltito* is more prevalent. In like manner, the long forms of *ebrio*, *momia*, *sepia*, and *vatio* are preferred in the US, while the rest of the countries either prefer the short forms or demonstrate no statistical preference. Only two participants chose the long form of all of the test words, while only 90 speakers (14%) preferred short forms for all test words ending in *-io/a*.

Word	AR	BO	CL	CO	CR	CU	DO	EC	ES	GT	HN	MX	PA	PE	PR	SA	US	UY	VE
<i>bestia</i>		s		S	S	S	S				s		s			s			
<i>cilia</i>					S					S	s		s				L		
<i>cumbia</i>		s	S	S	S				L				s			s			
<i>ebrio</i>				S	S	S			S	S			s			s	L	S	S
<i>fluvio</i>		s	S	S												s			S
<i>folio</i>				S	S	S		S		S			s			s		S	
<i>hostia</i>		s			S	L					s		s			s			S
<i>juicio</i>				L									l						
<i>lacio</i>				L			L								L				L
<i>lidia</i>				L					L	l			l						
<i>limpio</i>	S		L	L		L	L		S			L	l	L	L		L	S	L
<i>mafia</i>			L		L	L				L	l		l		L			L	L
<i>magia</i>	L	l		L	L						l		l		L	l		L	
<i>momia</i>		l		L					L		l		l				S		L
<i>napia</i>				L	L				L				l			l			
<i>necio</i>				L	L	L			L	L			l					L	
<i>norio</i>			S		L				L		s		l						
<i>novio</i>									S										
<i>nutria</i>	L	l			L								l			l			L
<i>ocio</i>		l	L	L	L								l			l			L
<i>paria</i>					L			S	L				l			l			
<i>patria</i>		l	L		L	L							l			l			L
<i>podio</i>				L	L								l			l			L
<i>rabia</i>				L	L		L			L	l		l	L	L	l			L
<i>regio</i>				L	L					L	l		l		L	l		L	
<i>rubio</i>	L		L	L	L		S		S	L		L	l	L	S	l		L	L
<i>rucio</i>		l	L		L			S					l			l		L	
<i>sabio</i>		l	L		L			L		L	l		l		L	l			
<i>sepia</i>	S	s		S	S	S							s			s	L		
<i>simio</i>	S	s	S		S	S					s		s		S	s			
<i>socio</i>			L		S				S	S			s			s			
<i>sodio</i>		s	L	S	S								s						
<i>vatio</i>		s		S	S		S						s				L		S
<i>vicio</i>	S	s		S	S	S							s	S					

voltio		s	S		S	L		S					s		s				
# of long	3	7	9	13	15	6	3	1	6	7	7	2	19	3	7	12	5	6	10
# of short	4	10	5	9	13	6	3	4	5	4	5	0	13	1	2	10	1	3	4

Results based on 1-9 participants are in lower case.

Table 4. Relationship between country and test items ending in *-io/a*.

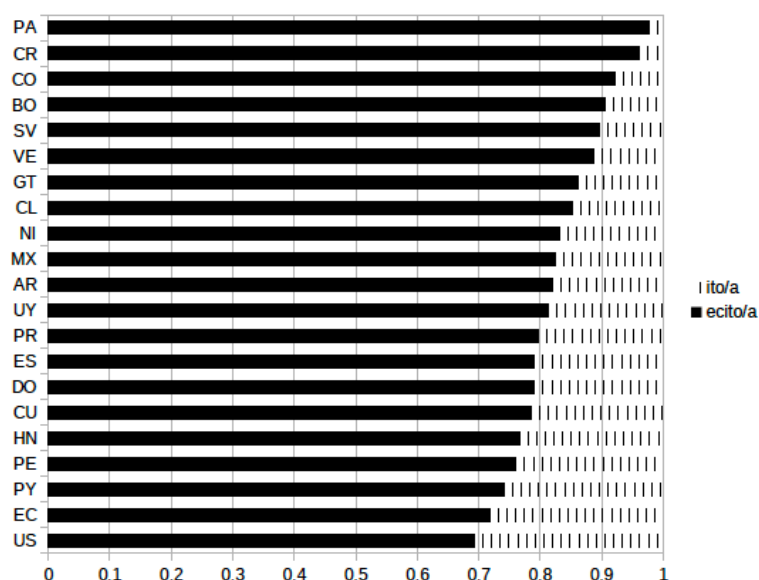


Figure 3. Proportion of long and short forms of words ending in *-io/a* by country.

3.4 Results for words ending in *-e*

The independent variables that reached significance in the analysis of words ending in *-e* were gender ($F(1, 14845) = 5.48, p < .019$) and the interaction of test item and country ($F(551, 14845) = 1.98, p < .0005$). Women preferred the long forms in 79.2% of the cases, and men 80.6% (coefficient = .268). While this difference may be statistically significant, it is small enough that it is not of much interest. The word by country data are summarized in Table 5. Long forms are most prevalent throughout the Spanish-speaking world, and most speakers prefer long forms for most words since 79.4% or the responses were long. However, one speaker preferred the long forms of all of these test words, while 40 (6%) preferred the short form in every case.

A few words are unique in certain countries because the short forms predominate, in contrast to most countries which either prefer the long form or neither the long nor the short form to a significant level. For example, *gentita* is preferred in Peru, and *mugrita* and *ñamito* in Mexico. In the US *buquito*, *ejito*, *enita*, *fasita*, *plieguito*, and *torpito* were preferred in contrast to all other countries. In fact, US Spanish uses more short forms of these words than all other countries (Figure 4).

Word	AR	BO	CL	CO	CR	CU	DO	EC	ES	GT	HN	MX	NI	PA	PE	PR	PY	SA	US	UY	VE
<i>buque</i>	L	l	L	L	L	L				L			l	x		L	l	l	S	L	L
<i>carne</i>	L	l	L		S		S	L	L	S	s	S	s		L	L	l	s	S	L	S
<i>cisne</i>		l				L		L				S	l	l			l		S	L	
<i>deje</i>		s	L						L	L		L	l	l		L	s	l			L
<i>diente</i>	S	l	L	L	S	L		L	L	L		S	l	l	L		l		S	S	L
<i>dulce</i>	L		L	L			S		L	S		L		l	L	L	l	s		L	S
<i>eje</i>	L			L	L			L	L	L	l		l	l	L		s	l	S		L
<i>ene</i>		l		L	L					L			l	l			l		S		
<i>eñe</i>		l															l				
<i>fase</i>		l	L		L	L	L		L				l	l			l	l	S		
<i>fauce</i>		l						L						l							
<i>fuella</i>			L	L			S				l			l			l	l	S		
<i>fuete</i>				L	S				L					l			l	l			L
<i>gafe</i>					L				L				l				s		S	L	
<i>gente</i>		l		L				L	L					l	S		l	l		L	L
<i>leche</i>	S	l	L	L	S	S	S	L	L		s	S	s		L	L	s	s		S	S
<i>leve</i>		l	L	L		L			L		l		l	l		L	s	l		L	L
<i>liebre</i>				L				L	L	L	l		l	l				l		L	
<i>miope</i>		l	L			S	S			L	l		l	l				l			L
<i>mugre</i>	L	l	L	L				L	L	L		S		l				l		L	
<i>nene</i>	S	l	S	L			S			S						S	s			S	L
<i>ñame</i>				L	L							S		l		L	x	l			L
<i>pliegue</i>													l	l			l	l	S	L	
<i>rifle</i>	L	l	L	L	L			L					x	l			l			L	L
<i>suave</i>	L		L	L	L		L			L	l		l	l		L	l	l	L	L	L
<i>timbre</i>	L	l		L				L		L			l	l		L	l			L	
<i>torpe</i>	L	l	L		L	L				L	l		l	l		L	x	l	S	L	
<i>ubre</i>	L	l	L	L	L				L		l		l	l			s	l		L	L
<i>yate</i>		l	L				L			L	l		l	l		L	l	l		L	L
# of long	10	19	16	18	10	6	3	11	14	12	9	2	17	23	5	11	16	17	1	16	14
# of short	3	1	1	0	4	2	6	0	0	3	2	6	2	0	1	1	7	3	11	3	3

Results based on 1-9 participants are in lower case. 'X' indicates there is no data for that cell.

Table 5. Relationship between country and test items ending in -e.

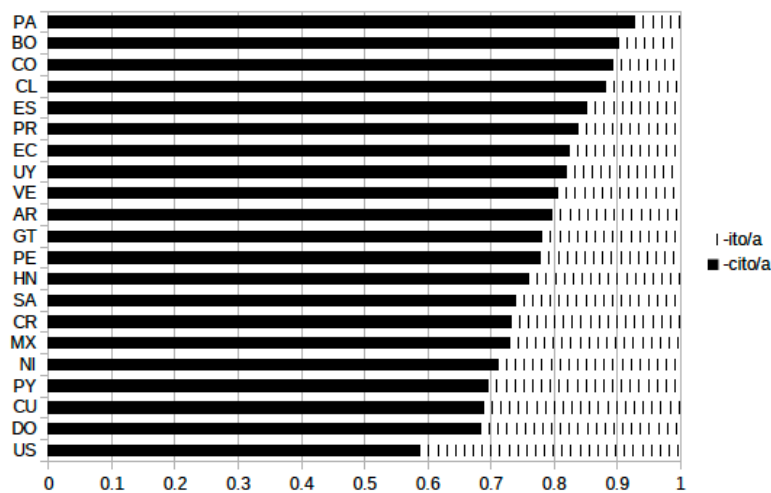


Figure 4. Proportion of long and short diminutive forms of words *ending in -e* by country.

4. Interim summary

The most salient finding of the present study is perhaps the least surprising: words have different diminutive forms depending on the country the speaker is from. More particularly, Spain stands out from the rest of the Spanish-speaking countries in its preference for the long diminutives of monosyllabic words and words with stem diphthongs (e.g. *piececito*, *trenecito* over *piecito*, *trencito*, and *viejecito*, *pueblecito* over *viejito*, *pueblito*). Individual variation is also apparent in the data. The majority of speakers preferred the long forms of most words ending in *-e* and *-io/a*. In contrast, few speakers preferred only short or long diminutives of all test words with stem diphthongs.

Some social variables were also influential. The long forms of diminutives of monosyllabic words were favored by older speakers and disfavored by younger speakers. This apparent time difference may signal a change in progress that may eventually eliminate long forms such as *piececito* and *trenecito*. Gender also arose as a significant factor. More than men, women preferred the long diminutives of words with stem diphthongs, and the short diminutives of words ending in *-e*. However, the differences, while statistically significant are so small that they are not very telling. This may be a case where a small difference can become statistically significant given enough data.

Spanish speakers in the US prefer *buquito*, *ejito*, *enita*, *fasita*, *plieguito*, and *torpito* in contrast to all other countries. If US Spanish were predominately influenced by one country, one would expect the results from the US to be more comparable to those of another country, but they do not. If US Spanish were merely a mixture of different varieties of Spanish, one would expect that the different influences would result in a combination of long and short forms, and yield primarily insignificant results for any given test word. Therefore, the most interesting finding is that US Spanish uses more short forms of these words than all other countries by a fairly wide margin (Figure 4).

5. Diminutive formation in Spain and Mexico

The previous sections document the factors that influence diminutive formation across the Spanish-speaking world. The fact that it is primarily dependent on individual words in a given country suggests that other factors may be relevant within a specific country, but are overridden by the strong country by test item difference. It is not feasible to examine all 21 countries, especially since a number of countries only have a handful of participants. Nevertheless, many participants were from Spain and Mexico (168 and 96, respectively), which makes statistical evaluation of these countries possible. Researchers who are interested in other country specific results may access the survey data online.⁸

5.1 Results for words ending in -e in Mexico

Diminutivization of words of this type was influenced by gender ($F(1, 2238) = 7.94$, $p < .005$), education ($F(2, 2238) = 7.84$, $p < .0005$), and test item ($F(28, 2238) = 15.26$, $p < .0005$). The coefficient of .631 indicates that women preferred short forms such as *lechita* and *carnita* slightly more than men. In Mexico, education is also related to diminutive formation. Figure 5 shows that the use of long forms is more prevalent among

⁸ URL will appear here in the printed version.

the more educated, while short forms are associated with lower educational levels. This is consistent with the data from all countries combined. The LSD post hoc test of these data reveals that participants with a college degree differ significantly from those with only a secondary or primary education, while participants with a primary or secondary education do not differ from each other ($p < .05$). The influence of individual words is shown in Figure 6. The LSD post hoc analysis of these data indicates which words deviate significantly from the overall mean. The words that are prefixed with ‘-’ are favored with the short allomorph, while words prefixed with ‘+’ favored the long allomorph more often.

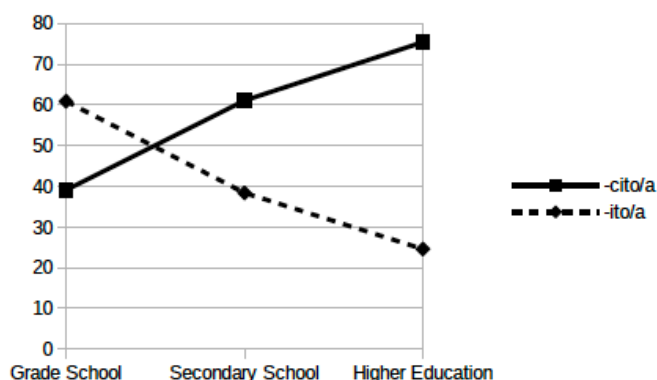


Figure 5. The influence of education on diminutive preference of words ending in -e in Mexico.

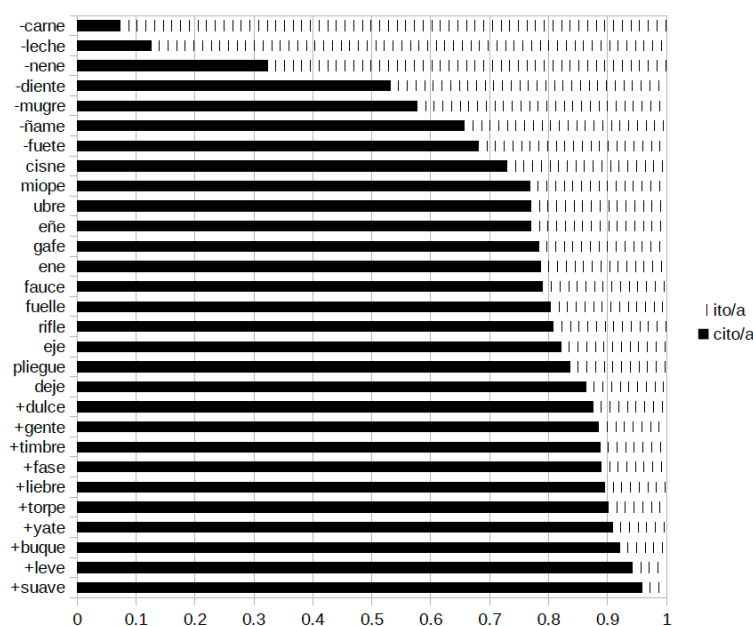


Figure 6. Proportion of long and short diminutive forms of test words ending in -e in Mexico.

5.2 Results for words with diphthongs in Mexico

The only significant variable on words with the diphthongs *-ie-*, *-ue-* in the stem was the particular word ($F(28, 2383) = 20.87, p < .0005$). The results of the LSD post hoc are summarized in Figure 7 where words where short forms were significantly favored more than the overall mean are marked with ‘-’, while words that were favored with long forms are marked with ‘+’.

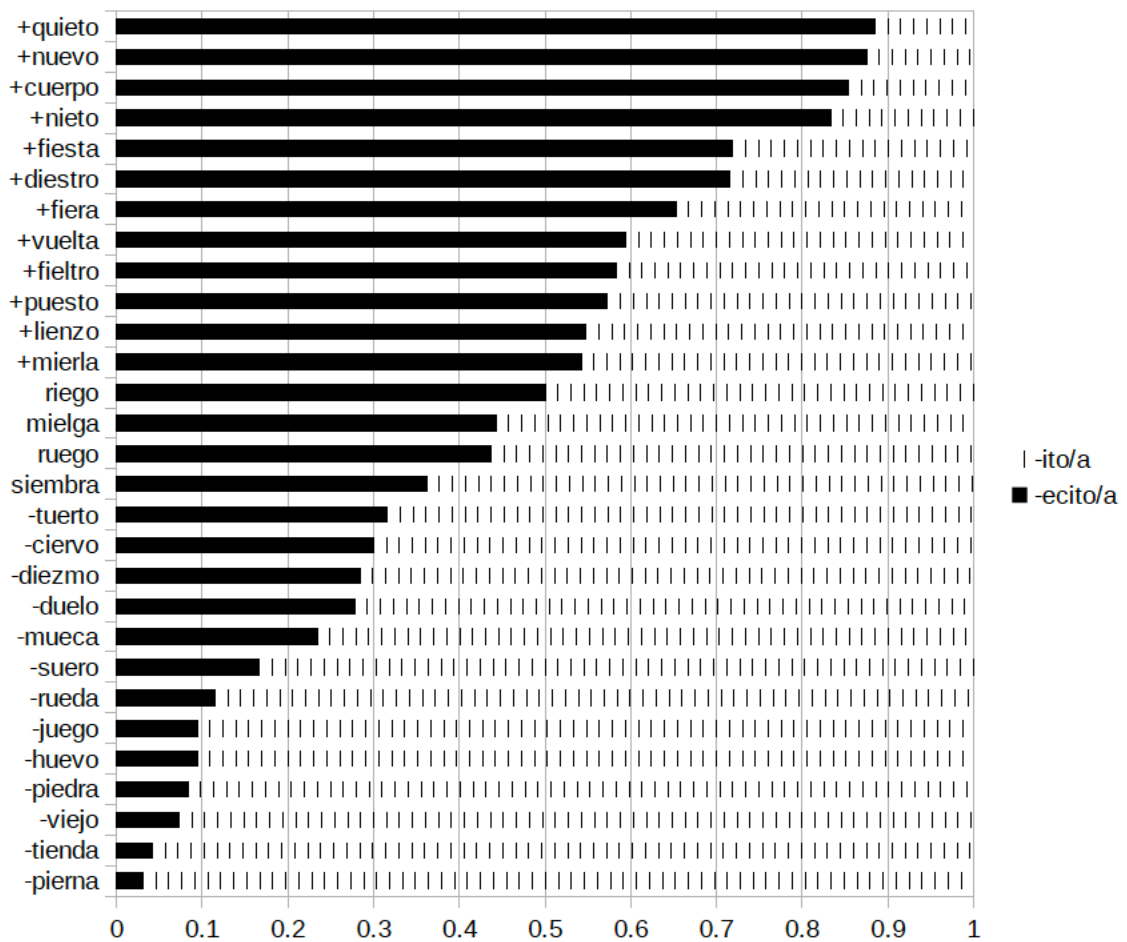


Figure 7. Proportion of long and short diminutive forms of test words with diphthongs in Mexico.

5.3 Results for words ending in *-io/a* in Mexico

Diminutive formation of words of this type was influenced by gender ($F(1, 2576) = 9.23, p < .002$), education ($F(2, 2576) = 6.91, p < .001$), and test item ($F(34, 2576) = 15.26, p < .0005$). The coefficient of $-.785$ for gender shows that women use fewer long forms than men. Men provided 55% long form while women only 45%. As far as education is concerned, the longer forms of the diminutives were favored by participants with more education. The LSD post hoc test shows significant differences between responses for participants with degrees in higher education compared to those with only a secondary or a primary school education ($p < .05$). There are no significant differences between participants with a primary or secondary school education (Figure 8). The results of the post hoc test for individual words is summarized in Figure 9. Words that are favored with short forms significantly more often than the overall mean are marked with '-', while words with significantly more long forms are marked with '+'.

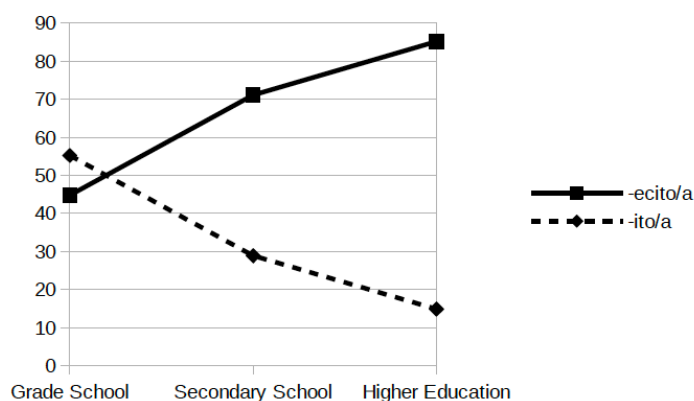


Figure 8. The influence of education on diminutive preference of words ending in *-io/a* in Mexico.

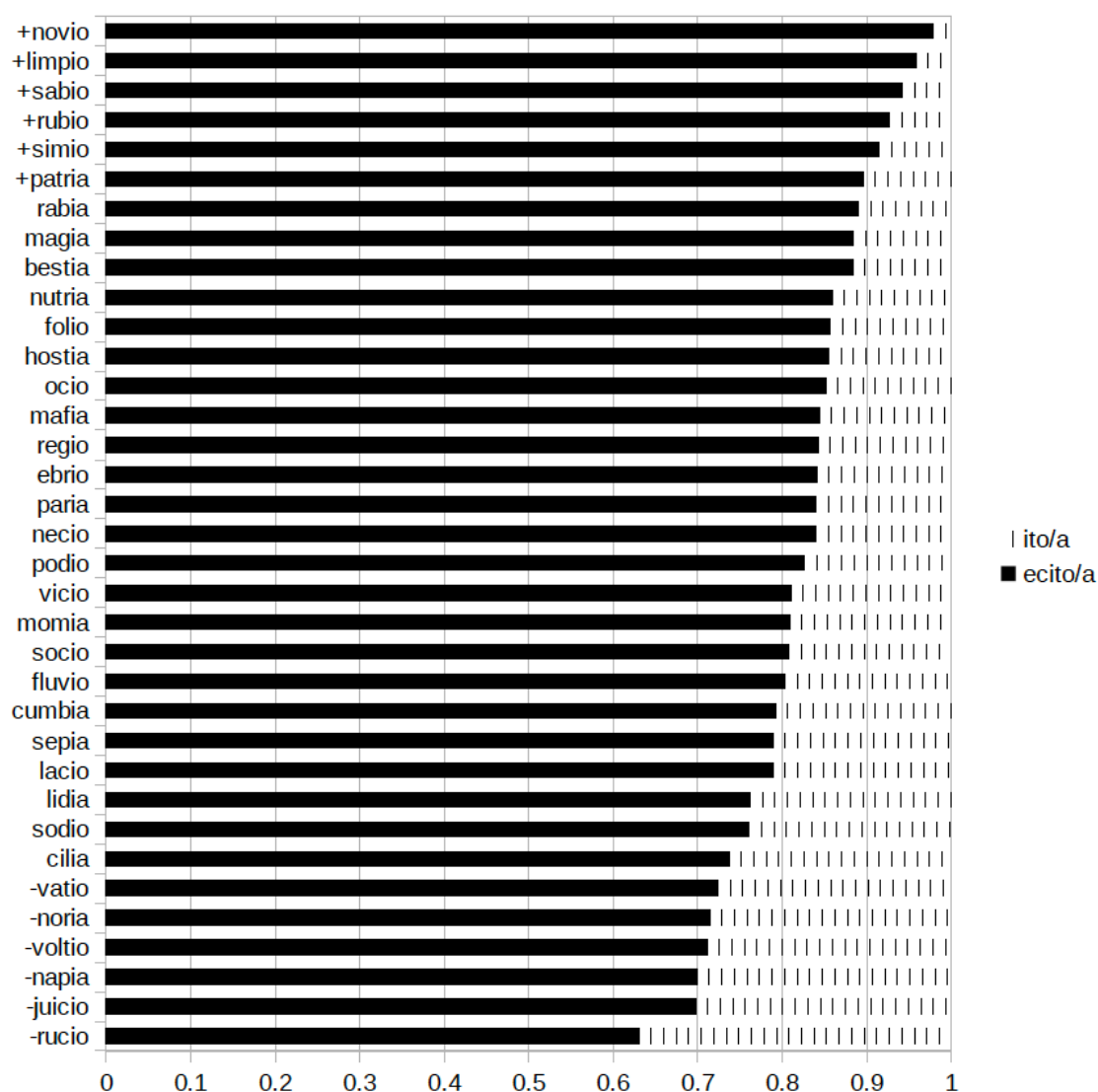


Figure 9. Proportion of long and short diminutive forms of test words ending in *-io/a* in Mexico.

5.4 Results for monosyllabic words in Mexico

Diminutive formation of monosyllabic was influenced by gender ($F(1, 378) = 4.24, p < .04$) and test item ($F(3, 378) = 32.74, p < .0005$). The coefficient of $-.639$ gender indicates that men favored the long diminutive forms more than women (55.6% men, 33.2% women). In the LSD post hoc analysis by test item (Figure 10), *trenecito* was preferred more than *trencito*, while *tecito* and *piecito* were favored over their long counterparts.

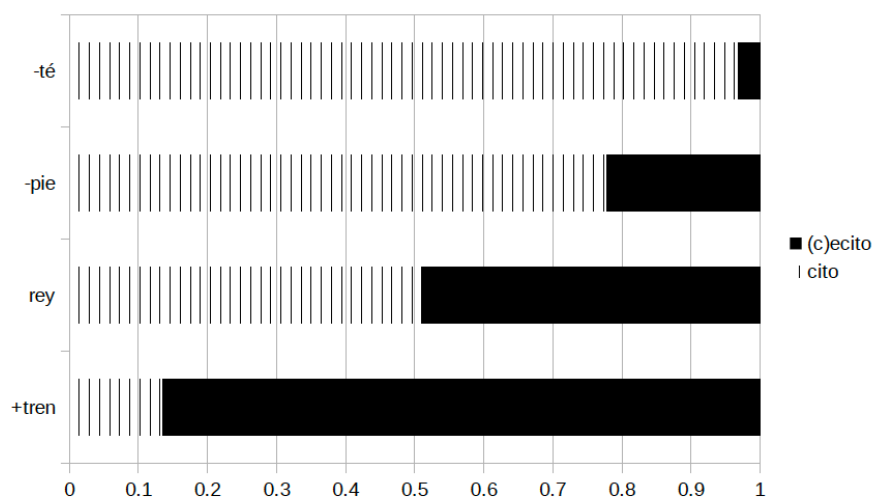


Figure 10. Proportion of long and short diminutive forms of monosyllabic words in Mexico.

5.5 Results for words ending in -e in Spain

The only significant factor for words of this type was test item ($F(28, 3933) = 15.00, p < .0005$). The results by word are summarized in Figure 11. Participants from Spain favored the long forms for most words, with *nenito* rather than *nenecito* as the primary exception.

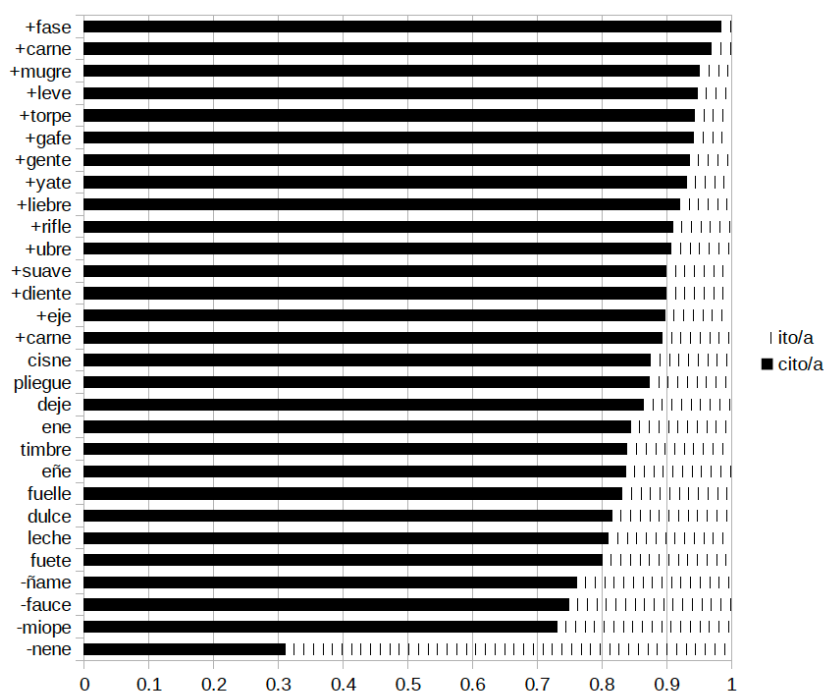


Figure 11. Proportion of long and short diminutive forms of words ending in -e in Spain.

5.6 Results for word containing -ie-, -ue- in Spain

Once again, the sole significant variable is test word ($F(28, 4247) = 14,74, p < .0005$). The result for individual test items appears in Figure 12.

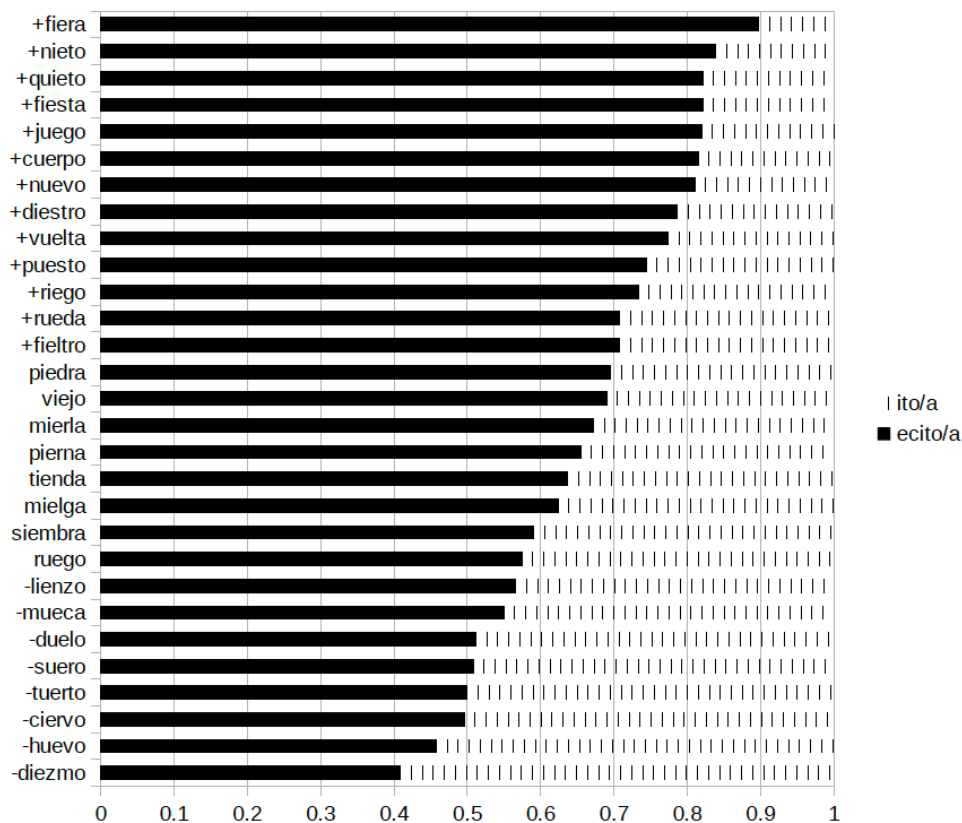


Figure 12. Proportion of long and short diminutive forms of words with stem diphthongs in Spain.

5.7 Results for words ending in -io/a in Spain

Once again, only test item is significant ($F(34, 4494) = 21.43, p < .0005$). As Figure 13 demonstrates, the long forms are preferred in the majority of cases. The most salient exceptions are *rubito* and *limpito*.

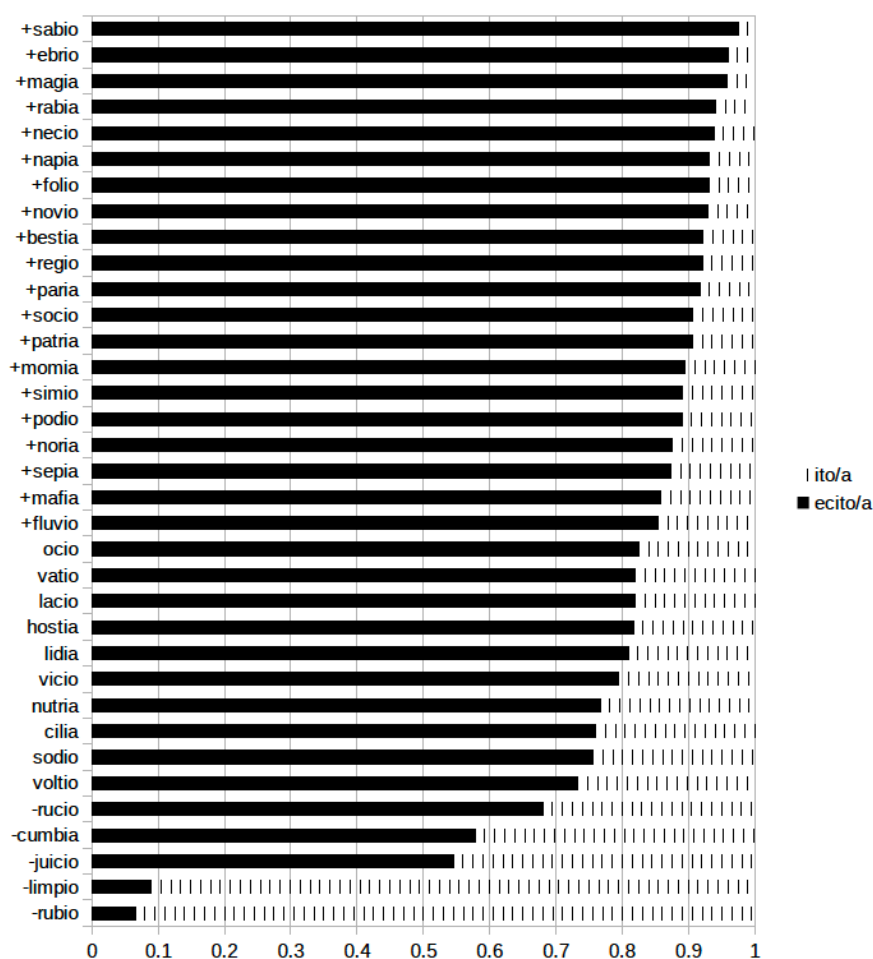


Figure 13. Proportion of long and short diminutive forms of words ending in *-io/a* in Spain.

5.8 Results for monosyllabic words in Spain

While it is not surprising that test item is a significant factor ($F(3, 659) = 49.62$, $p < .0005$), so are age ($F(3, 659) = 3.53$, $p < .015$) and education ($F(2, 659) = 6.06$, $p < .002$). The use of the long diminutive forms is associated with older speakers. The LSD post hoc analysis shows significant differences between the youngest group and the 50+ group as well as between youngest group and the participants in their 30s, but surprisingly not between the youngest participants and those in their 40s. As far as education is concerned, higher levels of education are associated with more use of the longer diminutives (Figure 14). The post hoc analysis shows that all educational levels differ from each other. The higher use of the long forms among older speakers and the more educated suggest that those forms carry more prestige but are falling out of use

by younger speakers. As far as the individual words are concerned (Figure 15), *tecito* was favored more often than *tececito*, while *piececito* and *trenecito* were preferred over *piecito* and *trencito*

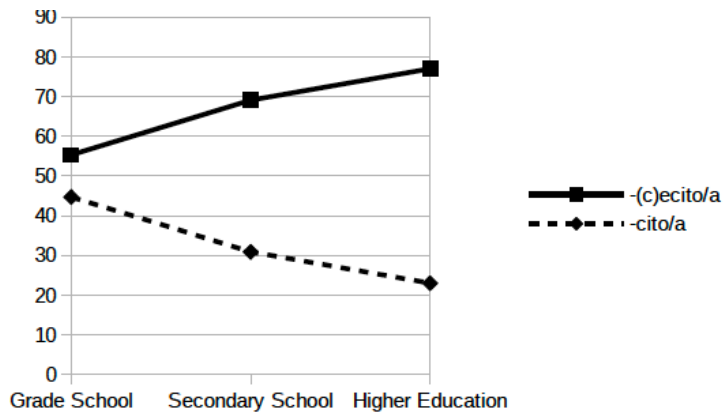


Figure 14. Influence of education on monosyllabic words in Spain.

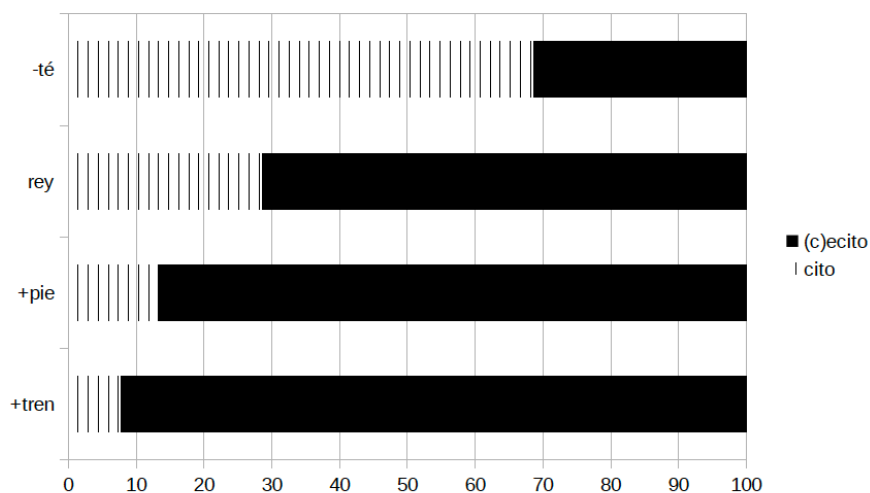


Figure 15. Percent of long and short diminutive forms of monosyllabic words in Spain.

6. Conclusions

Perhaps the most consistent result of the survey, and the least novel, is that diminutives vary a great deal from country to country. The second finding is that the diminutive form of a single word varies a great deal within a single country, which casts

doubt on any claim that the diminutive of word X in country Y is always Z. Instead, there are merely statistical tendencies for one diminutive form to be more prevalent than another in a particular country. However, in spite of all the variation evidenced in the survey, a number of trends were still found across all 21 Spanish-speaking countries for certain words. For example, the countries that do show a significant preference for one form over the other favor *suavecito*, *timbrecito*, *liebrecita*, and *riflecito* over *suavito*, *timbrito*, *liebrita*, and *riflito*. In like manner, *neciecito*, *rabiecita*, *simiecito*, and *viciecto* are much more likely diminutives than their shorter relatives *necito*, *rabita*, *simito*, and *vicito*.

The survey contained words that fall into four distinct classes, and the results of the survey clearly demonstrate that in no country are all of the diminutives in a single word class formed in the same fashion. Any proposed process of diminutivization that assumes they are is seriously flawed. The same thing is evidenced for individual speakers as well. Speakers who preferred either the short or long diminutives for all of the test words in any one of the four word classes are the exception not the norm. Most speakers preferred some mixture of long and short diminutives for words in a particular class.

Among the by results by country a few extremes are worthy of mention. The first is that Spain stands out when compared to the other 20 countries because it strongly prefers the long diminutives of all the monosyllabic words, as well as for most of the words with stem diphthongs (e.g. *nietecito*, *viejecito*). Argentina and Uruguay fall on the other end of the spectrum as far as the diminutives of words with stem diphthongs is concerned. Participants from these countries have a significant preference for the short forms of most of these test items (e.g. *nietito*, *viejito*), while the remainder of the other countries fall somewhere between these extremes.

The Spanish spoken in the US is also extreme in a number of ways. For words ending in *-io/a*, US speakers preferred short diminutives more than all the other countries. Consider *momia*, for example. Most countries demonstrate either no statistical preference between *momiecita* and *momita* or demonstrate a significant preference for *momiecita*. In the US, on the other hand, the clear winner is *momita*. Diminutive formation in the US goes against the trend for a number of other words as well. While most countries either prefer *cilita*, *vatito*, *ebrito*, and *sepita*, or are statistically ambivalent between the long and short diminutives, US Spanish prefers the long forms *ciliecita*,

vaticito, *ebriecito*, and *sepiecita* in spite of their overall preference for the short forms in general. In a similar fashion, US Spanish is more likely than any other country to prefer the short diminutives of words ending in *-e*. For example, only the US preferred the short forms *buquito*, *ejito*, *enita*, *fasita*, *plieguito*, and *torpito* over their long counterparts, in contrast to the trend in all other countries.

If diminutive formation in the US were heavily influenced by the Spanish of one country, then one would expect diminutives in the US to mirror those of some other country. The tables that summarize the country by test item results give no evidence of this. On the other hand, if Spanish in the US merely reflects the processes of diminutivization imported from many different countries, the expected outcome would be a mixed bag. That is, the trend in one direction from one country would be offset by the trend from another country in the opposite direction, and few significant differences would be expected. Instead, US participants demonstrate many significant preferences, some of which are either more extreme than, or contradict the general trends in the Spanish speaking world. The only reasonable conclusion to make is that US Spanish has developed its own system for producing diminutives apart from other Spanish speaking countries, something that further research should investigate.

One aspect of diminutive formation that has not been examined in previous studies are the social variables: education, age, and gender. The effect of some of these variables was observed with the monosyllabic test words; across all countries the long forms were preferred more among older participants. This same trend was also observed in Spain. This apparent time variation suggests that long monosyllabic diminutives may be dying out in favor of their short counterparts.

In addition, education level was also related to the preference for long diminutives with monosyllabic test items in Spain. As education level increased, so did the preference for the long forms. Education was also a significant predictor in Mexico, where preference for the long diminutives of words ending in *-e* and those ending in *-io/a* rose along with the level of education of the speakers. In Mexico gender was related to the short forms being favored. Specifically, Mexican women preferred the short diminutive for monosyllabic test items as well as for words ending in *-io/a*. This is in direct opposition to

the effect of education on words ending in *-io/a*. (The interaction between education and gender was tested and found to be insignificant.) If the long forms are favored more by the educated, that may suggest that they are associated with more prestigious speech, and that the speech of Mexican men trends toward the prestige diminutives more than women.

A good number of highly infrequent test items such as *simio* and *cilia* were included since the chances that any speaker had every produced or heard their diminutive would be quite low. The idea behind including them in the study was that highly frequency forms that have been heard and produced many times may be stored in the mental lexicon. Diminutivization for them could be a matter of lexical retrieval. In contrast, the diminutive form of a low frequency word would need to be uniquely derived by some online process. However, since frequency was not a significant variable, this hypothesis was not supported. The present study included only a small number of high frequency test items, and further research into diminutive formation that tests a larger number of words is clearly warranted.

Like most research, rather than answering lots of questions, it answers a few and raises even more. Specifically, what phonetic or other linguistic qualities of particular words influence them to take long versus short diminutives? Why is the diminutive of *bestia* *bestita* more often than *bestiecita*, while *rabiecita* is more common than *rabita* as the diminutive of *rabia*? Are diminutives conditioned by the speech context? One participant commented that *miopito* would be his general preference unless he were mocking a sibling who now had to wear glasses. In that context he would use *miopecito* instead.

A final note concerns methodology. Invitations were sent via Facebook, and participants were encouraged to pass the link along to other Spanish speakers. In addition, a task description was placed on Mechanical Turk individually for each country. In spite of this, fewer than ten speakers completed the survey in six countries, while there were about 100 or more participants in three countries. In fact, the surveys in those countries with many participants were closed early to prevent an even more lopsided distribution of the data. Mechanical Turk is an impressive tool for gathering

data, but falls short in one regard; it does not attract participants from all Spanish-speaking countries equally.

References

- AMBADIANG, Théophile (1996) "La formación de diminutivos en español: ¿Fonología o morfología?", *Lingüística Española Actual*, 18, 175-211.
- AMBADIANG, Théophile (1997) "Las bases morfológicos de la formación de diminutivos en español", *Verba*, 24, 99-132.
- BERMÚDEZ-OTERO, Ricardo (2007) "Spanish pseudoplurals: Phonological cues in the acquisition of a syntax-morphology mismatch", *Proceedings of the British Academy*, 145, 231-269.
- BERMÚDEZ-OTERO, Ricardo (2013) "The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features", *Probus*, 25, 3-103.
- BRADLEY, Travis G. & Jason SMITH (2011) "The Phonology-morphology Interface in Judeo-Spanish Diminutive Formation: A Lexical Ordering and Subcategorization Approach", *Studies in Hispanic and Lusophone Linguistics*, 4, 247-300.
- CALLEBAUT, Sien (2011) *Entre sistematización y variación. El sufijo diminutivo en España y Hispanoamérica*, Masters Thesis, University of Ghent.
- CASTILLO VALENZUELA, Rosario & Rosa María ORTIZ CISCOMANI (2013) "Diminutivo y aspecto nominal en español", *Revista de Humanidades*, 27, 155-172.
- CASTRO, Obdulia (1998) "La formación del diminutivo en español y en gallego: Procesos morfológicos simples; implicaciones teóricas complejas", in Andrés Acosta Félix, Zarina Estrada Fernández, Max Figueroa Estava & Gerardo López Cruz (ed.), *IV Encuentro internacional de lingüística en el noroeste*, Sonora, México: Universidad de Sonora, 135-159.
- COLINA, Sonia (2003) "Diminutives in Spanish: A Morpho-phonological Account", *Southwest Journal of Linguistics*, 22, 45-88.
- CROWHURST, Megan J. (1992) "Diminutives and Augmentatives in Mexican Spanish: A Prosodic Analysis", *Phonology*, 9, 221-253.
- EDDINGTON, David (2002) "Spanish Diminutive Formation without Rules or Constraints", *Linguistics*, 40, 395-419.

- ELORDIETA, Gorka & María M. CARREIRAS (1996) "An Optimality Theoretic Analysis of Spanish Diminutives", in Lise M. Dobrin, Kora Singer & Lisa McNair (ed.), *Proceedings from the Main Session of the Chicago Linguistic Society's 32nd Meeting*, Chicago: Chicago Linguistic Society, 49-60.
- FONTANELLA, Maria Beatriz (1962) "Algunas observaciones sobre el diminutivo en Bogotá", *Thesaurus: Boletín del Instituto Caro y Cuervo, Bogotá*, 17, 556-573.
- GAARDER, Bruce A. (1966) "Los llamados diminutivos y aumentativos en el español de México", *Publications of the Modern Language Association of America*, 81, 585-595.
- HORCAJADA, Bautista (1988) "Morfología de los diminutivos formados sobre bases consonánticos monosílabas", *Filología Románica*, 5, 55-72.
- JAEGGLI, Osvaldo A. (1980) "Spanish Diminutives", in Frank Nuessel Jr. (ed.), *Contemporary Studies in Romance Languages: Eighth Annual Linguistics Symposium on Romance Languages*, Bloomington, IN: Indiana University Linguistics Club, 145-158.
- MIRANDA, Inés Miranda (1999) *An Optimality Theoretic Analysis of Nicaraguan Spanish Diminutivization: Results of a Field Survey*, Dissertation, University of Washington.
- NORRMANN-VIGIL, Ingrid (2012) "Accounting for Variation of Diminutive Formation in Porteño Spanish", *Mester*, 41, 99-122.
- PRIETO, Pilar (1992) "Morphophonology of the Spanish Diminutive Formation: A Case for Prosodic Sensitivity", *Hispanic Linguistics*, 5, 169-205.
- REAL ACADEMIA ESPAÑOLA (2009) *Nueva Gramática de la lengua española*, vol. 1. Madrid: Espasa.
- REYNOSO NOVERÓN, Jeanett (2005) "Procesos de gramaticalización por subjetivación: el uso del diminutivo en español", in David Eddington (ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, Somerville: Cascadia Proceedings Projects, 79-86.
- ROJAS, Nelson (1977) "Aspectos de la morfonología del diminutivo -ito", in M. Chevalier et al. (ed.), *Actas del Quinto Congreso Internacional de Hispanistas*, vol. 2, Bourdeaux: Instituto de Estudios Ibéricos e Iberoamericanos, 743-751.
- RUIZ DE MENDOZA IBÁÑEZ, Francisco José (1996) "Some Notes on the Grammatical of the Spanish -ito/-illo Diminutives and Their Translation into English", *Pragmalingüística* 3-4, 155-172.
- SMITH, Jason Allen (2011) *Subcategorization and Optimality Theory: The Case of Spanish Diminutives*, Dissertation, University of California, Davis.
- STEPHENSON, Tamina (2004) "Declensional-type Classes in Derivational Morphology: Spanish Diminutives Revisited", Manuscript, MIT.

ZULUAGA OSPINA, Alberto (1993) “La función del diminutivo en español”, in Rubén Paez Patino (ed.),
Thesaurus: Boletín del Instituto Caro y Cuervo Muestra Antológica 1945-1985, vol. 1, Santa
Fe de Bogotá: Instituto Caro y Cuervo, 305-330.

Appendix 1: Test items

-io/a words

bestia	juicio	napia	patria	sepia
cilia	lacio	necio	podio	simio
cumbia	lidia	noria	rabia	socio
ebrio	*limpio	*novio	regio	sodio
fluvio	mafia	nutria	*rubio	vatio
folio	magia	ocio	rucio	vicio
hostia	momia	paria	sabio	voltio

-ie-, -ue- words

ciervo	fiera	mierla	*puesto	suero
*cuerpo	*fiesta	mueca	*quieto	*tienda
diestro	*huevo	*nieto	riego	tuerto
diezmo	*juego	*nuevo	*rueda	*viejo
duelo	lienzo	*piedra	ruego	*vuelta
fieltro	mielga	*pierna	siembra	

-e words

buque	eje	fuete	miope	*suave
carne	ene	gafe	mugre	timbre
cisne	eñe	*gente	ñame	torpe
deje	fase	*leche	*nene	ubre
*diente	fauce	leve	pliegue	yate
*dulce	fuella	liebre	rifle	

Monosyllabic**words**

*pie

*rey

*té

*tren

Truth testers

*árbol

*canción

