¿Vas pa(ra) Málaga? The Reduction of *para* in Málaga, Spain: Effects of Frequency, Syntactic Category, and Social Factors

Rob Bedinghaus

1. Introduction

Phonetic reduction in spoken discourse has been attributed to various causes, including sociolinguistic factors (Bentivoglio, Guirado & Suárez, 2005; Guirado, 2007) and frequency of usage in discourse (Bybee & Scheibman, 1999; Bybee, 2002a, 2002b; Díaz-Campos, Fafulas & Gradoville, 2012; Díaz-Campos & Gradoville, 2011; Díaz-Campos & Ruiz-Sánchez, 2008; File-Muriel, 2007, 2009; Gradoville, 2012; Johnson, 1997; Pierrehumbert, 2001; Rao, 2009). Bybee (2002, 2006) argued for the *reducing effect* of frequency in which high frequency of use in discourse leads to greater phonetic reduction. She states, "This model predicts that the frequency with which words are used in the contexts for change will affect how readily the word undergoes a change in progress" (2002, p. 261). The argument is that sound change is gradual, not abrupt, and that it spreads from word-to-word due to connections between similar exemplars stored in memory. Elsewhere Bybee argues that, "[the] articulatory representation of words and sequences of words is made up of neuromotor routines. When sequences of neuromotor routines are repeated, their execution becomes more fluent" (2006, p. 715). In other words, the accumulation of exemplars in the memory and the repetition of these in speech cause a more fluent production that produces phonetic reduction and subsequently causes lexical diffusion of a sound change from more frequent words to less frequent words over time. Bybee provides evidence from /t/-/d/ deletion in American English and intervocalic /d/ deletion in New Mexican Spanish that higher frequency words exhibited a higher rate of deletion.

Bybee and Scheibman (1999) argued that the reducing effect of frequency applies not only to individual lexical items, but also to words that are frequently used together in discourse. They argue that such word groupings (henceforth *collocations*) are stored together in the lexicon and tend to reduce phonetically when there is a tighter bond between them caused by frequency of use. The theory is supported by data in which they demonstrate that *don't* in American English reduces more often in highly frequent collocations such as *I don't know* (e.g., a flap and schwa as in [aj.ɾə.no]), and less often in lower frequency collocations such as *we don't think* [wi.doʊnt.θɪŋk]. Similarly, Jurafsky et al. (2001) proposed a probabilistic model that investigates not only the effect of relative frequency of a lexical item, but also the effect of predictability on phonetic reduction (Probabilistic Reduction Hypothesis). This model, similarly to Bybee and Scheibman (1999), claims that words are very predictable from their neighboring words, and that more predictable or probable words and collocations are more likely to be reduced. They further claim, with Bybee and others, that these probabilistic relations between words are represented and stored in the speaker's mind. Their hypothesis is supported by corpus data that show that more probable words (i.e., relative higher token frequency in the corpus) undergo more reduction. However, they also found a discrepancy regarding the effect of conditional probability (i.e., the probability of a target word appearing given the preceding or following word) on the reduction of content versus function words. Interestingly, predictability from neighboring words played a strong role in the reduction of high-frequency function words, but had a

significantly weaker effect on the reduction of high-frequency content words. Shi et al. (2005) also argued that theories of lexical representation need not only account for frequency, but also for syntactic category (i.e. content vs. function words). They found that frequency and syntactic category contribute differently to different phonological processes in English. These claims regarding frequency and syntactic category have yet to be investigated for the reduction of *para* in Spanish.

The purpose of the current study is to test Bybee and Scheibman's theory of collocation frequency effects on the reduction of the Spanish preposition *para* (/paɾa/, 'for') to [pa] in Málaga Spanish, as well as the effects of syntactic category (content vs. function words following *para*) and the social factors of age and sex. It is well known that *para* is variable throughout the Spanish-speaking world. However, few empirical studies have documented it. Bentivoglio et al. (2005) and Guirado (2007) studied *para* reduction in Caracas, Maracaibo, and Mérida, Venezuela, and found that it is conditioned by various social and linguistic factors, such as the semantic meaning of *para*, whether a consonant or vowel followed *para*, age, socioeconomic class, and dialect. More recently, Díaz-Campos et al. (2012) examined the effects of collocation frequency (i.e., frequency of the collocation of *para* with a preceding or following word) on the reduction of *para* in Caracas, Venezuela. However, no study to date has investigated frequency effects, social factors, and syntactic category in one cohesive study. Furthermore, studies to date on the reduction of *para* have been focused solely on Venezuela. Thus, the current study contributes to our understanding of the effects of frequency, syntactic category, and social factors on phonetic reduction, and specifically the reduction of *para* in a variety of Spanish outside of Venezuela.

2. Previous literature
2.1. Málaga Spanish

Málaga is located in south central Andalusia in Spain. Andalusia is known for its linguistic innovation, particularly in more western cities like Seville. However, Villena Ponsoda (2008) describes the Spanish of Málaga as being in the process of slowly converging toward the Castillian standard over the last 60 to 70 years. He gives evidence from previous work that Málaga does not tend toward certain innovations that are found in more western varieties in Seville and Jerez, such as seseo, ceceo, and the de-affrication of /tʃ/ to [ʃ]. However, that does not mean that Málaga does not participate in any innovative phonological processes. Vida (2005) showed that Málaga, while it was not the *most* innovative dialect, was among the dialects with the highest rates of /s/-deletion among the dialects compared. Samper Padilla (2011) also showed that Málaga is among the more innovative dialects in the literature for the deletion of intervocalic /d/. It is because of these tendencies toward segmental reduction that the current study targets the reduction of the preposition *para* in Málaga.

2.2. The reduction of para in Spanish

Spanish dialectologists (Calcaño, 1949; Cuervo, 1939, 1944; Henríquez Ureña, 1940) have commented on the existence of *para* reduction, but until recently it was not well represented in empirical investigations. Until the more recent variationist work, the *pa-para* alternation was assumed to be in free variation. The first modern variationist work on *para* reduction comes from Bentivoglio et al. (2005) and Guirado (2007), who worked with the *Corpus Sociolingüístico de Caracas* (Bentivoglio & Sedano, 1993). Guirado (2007) also included corpora from Maracaibo, Venezuela (Chela-Flores & Gelmán, 1988) and Mérida, Venezuela (Dominguez, 1996). Both studies examined linguistic and social factors. Linguistic factors included whether the segment following *para* was a consonant or a vowel, and the semantic meaning of the preposition. The categories of meaning were directionality (e.g., voy *para* Caracas "I'm going *to* Caracas"), temporality (e.g., Somos amigos *para* siempre "We are friends *for*ever"), goal (e.g., Estudiamos *para* sacar buenas notas "We study *in order to* get good grades"), and particularity of the speaker (only in Guirado, 2007 - e.g., *Para* mí la amistad es importante "*To/for me* friendship is important"). Social factors included age (30-45 years, 60+ years), sex, and socioeconomic class (low, middle, high). Using a Varbrul analysis, both studies determined that socioeconomic class was the most significant factor with the lower class favoring reduction the most. Also, both studies found the semantic meaning of *para* to be the second-most significant factor,

with a directional meaning strongly favoring reduction and all other meanings disfavoring it. A following consonant favored and a following vowel disfavored reduction, and older speakers favored reduction while younger speakers disfavored reduction (see Díaz-Campos et al., 2011 for a more detailed analysis of social factors in the same Caracas corpus). Guirado (2007) also compared dialects and found that there was variation between the three dialects of Venezuelan Spanish studied. Caracas slightly favored reduction, while Mérida slightly disfavored it and Maracaibo disfavored reduction the most.

Díaz-Campos et al. (2012) is the only study to date to examine frequency effects on the *pa/para* alternation in Spanish. Also using the *Corpus Sociolingüístico de Caracas* (Bentivoglio & Sedano 1993), they extracted all cases of full and reduced *para* from the corpus, excluding certain cases that exhibited a pause preceding or following *para* and cases where two tokens of *para* were adjacent. They coded only for linguistic factors, which are the following: the manner and place of articulation of the segment following *para*, the semantic meaning of *para* (i.e., directional, goal, temporal, and other), the grammatical category of the word following *para*, and the frequency of the collocations WORD+*para* and *para*+WORD (i.e., the number of times a given word preceded and/or followed *pa/para* in the corpus). During the analysis, they decided to exclude the phonological and grammatical category factors due to interactions with other factors and for other reasons. They used a Varbrul analysis to determine the linguistic factors that most favored the reduction of *para.*

Their results showed that the factor with the most significant effect on the reduction of *para* was the meaning of the preposition. When *para* had a directional meaning, it was reduced most often, followed by a goal meaning, temporal meaning, and lastly all other meanings. The second most significant factor was the frequency of *para*+WORD collocations. The highest frequency collocations favored reduction and the second frequency group neither favored nor disfavored it. The lowest two frequency groups both disfavored reduction. The frequency of WORD+*para* collocations was the final significant factor, showing a very slight favoring of reduction with the high frequency collocations and very slight disfavoring with the low frequency collocations[1]. They concluded from the frequency analysis that higher collocation frequency indeed influenced the reduction of *para*, and that the effect was strongest for *para*+WORD collocations. They showed that the *para*+WORD category was made up of a smaller number of words that are each more frequently collocated with *para*, whereas the WORD+*para* category was made up of a greater number of words that are each less frequently collocated. It follows that the *para*+WORD collocations should have a greater influence on the data than WORD+*para* collocations. They further supported their argument by demonstrating that the nine most frequent words following *para* in the corpus represented two-fifths of the total data, and when removed, caused a significant decline in reduction in the *para*+WORD category (i.e., 51.1% to 10.7%), showing that phonetic context effects did not apply throughout the data. Furthermore, most of these words were function words as opposed to content words, an interesting fact that they did not discuss. Díaz-Campos et al. (2012) also noted that the frequency effect was gradient, showing the most reduction in the highest frequency collocations and decreasing with each lower frequency group, which supports Bybee and Scheibman (1999)'s claim that frequency effects are in direct correlation with the strength of collocations. Gradoville (2012) also found a gradient effect of frequency on the reduction of *para* in Carioca (Rio de Janeiro) Portuguese.

In conclusion, the previous literature has shown that various social, linguistic and frequency factors condition the reduction of *para*. However, the previous studies only represent Venezuelan Spanish. To date, no study has examined the role of frequency effects and sociolinguistic factors *together* on the reduction of *para* and in a non-Venezuelan dialect. Additionally, a discussion of the effects of content versus function words on *para* reduction is warranted. The current study begins to fill these gaps in the literature.

The current study addresses the following research questions:

1) Are there significant effects for frequency on the reduction of *para* in Málaga? If so, how do the results compare to Diaz-Campos et al. (2012)?

---

[1] There were four frequency groups for the para+WORD constituents and two for the WORD+para constituents.

2) Is there a discrepancy between *para*+function and *para*+content word collocations on the reduction of *para*? Does this analysis support the claims of Jurafsky et al. (2001) and Shi et al. (2005)?

3) What are the trends for the social factors of *age* and *sex* regarding the reduction of *para*? How do these results compare to previous sociolinguistic work by Bentivoglio, Guirado and Suárez (2005) and Guirado (2007) on Venezuelan Spanish?

3. Methodology

The current study replicates the methodology of Díaz-Campos et al. (2012), including all factors presented in their study and adding the social factors of age and sex of the speaker, as well as syntactic category (post-hoc). The data come from the Málaga corpus (Villena Ponsoda et al., 2002) that is part of the large-scale, multi-dialectal *Proyecto Para el Estudio Sociolingüístico del Español de España y de América* (PRESEEA). This study included data from 60-minute interviews with 23 speakers (11 male, 12 female) that represented three age groups (20-34, 35-54, 55 and above). The interviews discussed many different topics about the lives of the speakers, such as their history in Málaga, their family and work lives, politics, education, and more. All speakers were from Málaga and had a high level of education ('instrucción superior'), meaning at least university level. Their professions included professors (the majority), executives, engineers, doctors, teachers, students, homemakers, and administrative positions. Given that only the upper class data were available for the current study, socioeconomic class is not included as a factor in the analysis.

All instances of *para*, both full and reduced, were extracted from the corpus transcriptions done by Villena Ponsoda et al. (2002) using the search feature in Adobe Acrobat Pro. The reduction of *para* was marked in the transcriptions as *pa<(r)><(a)>* and was easily searched. Each transcription was originally done by one PRESEEA researcher, and subsequently revised two or three more times by different researchers each time. In this and other ways, the PRESEEA project is well-controlled and seeks to employ the same conventions for the many researchers involved across many dialects of Spanish. For these reasons, the transcriptions of the Málaga corpus transcribers were trusted for this study and not re-analyzed for accuracy. Also, all tokens of *para* with a preceding or following pause, as marked in the transcription, and all *para* doubles (i.e., para para) were excluded from the analysis in order to target fluid speech that was not interrupted by hesitations or repetitions. Each instance of the preposition was extracted with surrounding context and coded for the following linguistic and social factors:

1. The phonological status of the segment immediately following *para*:
   a. Height, frontness/backness of the following vowel
   b. Manner and place of articulation of the following consonant
2. The frequency of the collocations consisting of the word immediately preceding or following *para*[2]:
   a. WORD+*para* frequency (low = 1-3 tokens, high = 4-25 tokens)
   b. *para*+WORD frequency (low = 1-15 tokens, high = 16-97 tokens)
3. The meaning associated with the preposition:
   a. Directional
   b. Goal
   c. Temporal
   d. Other
4. Grammatical category of the following word (i.e., adverb, noun, article, verb etc.)
5. The age of the speaker (i.e., 20-34, 35-54, 55+)
6. The sex of the speaker

---

[2] The division of high and low frequency groups is arbitrary because of making a scalar variable categorical for use in Goldvarb. Diaz-Campos et al. (2012) recognize the same issue. The division here was made by looking at the distribution of the collocation frequency data and finding a representative cutoff.

A Varbrul analysis was conducted using Goldvarb X (Sankoff et al., 2005). The output of the binomial logistic regression gives a factor weight between 0 and 1 for each level of each factor. A factor weight above .50 indicates a favoring relationship between that factor and the reduction of *para*. A factor weight below .50 means that the factor does not favor reduction. Furthermore, the higher or lower the factor weight than .50 indicates a stronger positive or negative relationship between a given factor and the reduction of *para*.

4. Results

795 tokens of full and reduced *para* were coded and analyzed. Table 1 displays the overall distribution of full and reduced forms in the data. Of the 795 tokens, 39.4 percent (313/795) were reduced and 60.6 percent (482/795) full, a comparable figure to the 38.5 percent reduction observed in Díaz-Campos et al. (2012) and less than 51 percent in Bentivoglio, Guirado & Suárez (2005) and 48 percent in Guirado (2007). It is important to note, though, that Díaz-Campos et al. (2012) included the entire Caracas corpus, while Bentivoglio et al. (2005) and Guirado (2007) only analyzed part of that corpus. Consequently, Díaz-Campos et al. (2012)'s result is more indicative of *para* reduction in the Caracas corpus.

Table 1. Distribution of variants of *para* in the Málaga corpus of 23 speakers

	Reduced	**Full**	**N**
Tokens	313	482	795
Percentage	39.4%	60.6%	100%

4.1. Linguistic Factors

Examining the factors of the phonological status of the segment following *para* (Figure 1), reduction occurred most often when *para* was followed by a low vowel (44.1%, 52/118, e.g., [pa] [a]cá "to here"). There was 42.3 percent (143/338) reduction with a following obstruent (e.g., [pa] [t]i "for you"), 38.3 percent (49/128) with a following mid vowel (e.g., [pa] [e]studiar "in order to study"), 35.1 percent (13/37) with a following high vowel (e.g., [pa] [i]r "in order to go", and 32.2 percent (56/174) with a following sonorant (e.g., [pa] [n]osotros "for us"). Díaz-Campos et al. (2012) noted that there did not appear to be clear phonological patterning since the most favoring categories (obstruent and low vowel) are on the opposite sides of the scale typically used to classify the sonority of segments (Kenstowicz, 1994). The same is true here. Also, in the current study, there is not a clear advantage for having a consonant or vowel following *para* (40 percent reduction with a following vowel and 39 percent with a consonant).

Regarding the place of articulation of the following segment (Figure 2), *para* reduced most often when followed by a palatal (60%, 3/5, e.g., [pa] [j]evarlo "in order to carry it") or dorsal consonant (49%, 79/162, e.g., [pa] [k]e "so that"). However, it is important to note that there were only five cases of a palatal consonant following *para* in the entire dataset. With so few tokens, such an apparently strong effect is unreliable. Also, 58 percent (94/162) of all dorsal consonants were from the conjunction *que "that"*. Of the 79 reduced cases of *para* when followed by a dorsal consonant, 49 occurred when *para* was followed by *que*. Furthermore, only six different words made up 53.8 percent of all tokens following *para* that began with a back vowel, and of the 77 combined tokens of these six different words, 54.5 percent (42/77) were reduced. In contrast, 36 different words made up the other 46.2 percent of back vowel tokens. For the low vowel category, 65.2 percent of low vowel tokens following *para* were one of the six mentioned above. Clearly, there is a strong interaction between the phonological factors and frequency, which will be discussed later.

Figure 1. (%) Rates of *para* reduction by manner of articulation of the following segment
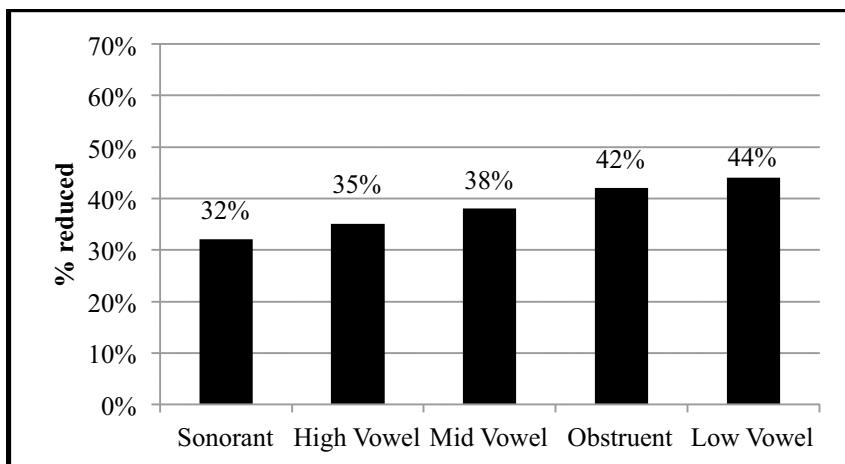
Figure 2. (%) Rates of *para* reduction by place of articulation of the following segment
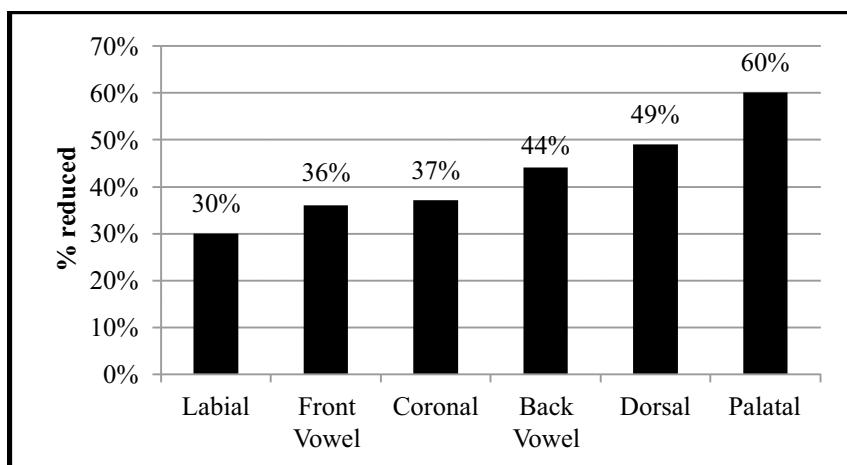
Looking at the results for grammatical category (Figure 3), it is clear that a following conjunction (e.g., mostly *que* 'that') coincided with *para* reduction the most, followed by adverbs (e.g., [pa] *allá* "to there") and articles[3] (e.g., [pa] *él* "for him"). The largest difference in reduction between two categories is a nine percent difference between articles and prepositions (e.g., [pa] *en* vez de… "for instead of…"). Importantly, it is not claimed here that grammatical category is the *cause* of reduction, but rather that there is an interaction between this factor and frequency that masks the frequency effects, which will be discussed later.

The results for reduction according to the meaning of the preposition *para* (Figure 4) show that *para* was reduced most often when it carried a directional meaning (e.g., [pa] *Málaga* "to Málaga"). A temporal meaning (e.g., [pa] *el mes siguiente* "for next month") was the second-most favoring of reduction. However, the temporal category consists of very few tokens relative to the other categories of prepositional meaning. Therefore, it is doubtful that this result is a reliable indicator of the reduction of *para* as a temporal preposition.

---

[3] Díaz-Campos et al. (2012) used the category *determiner*, whereas here *article* is used. The reason is that some determiners also fall under another category like adjectives (e.g., possessives and demonstratives).
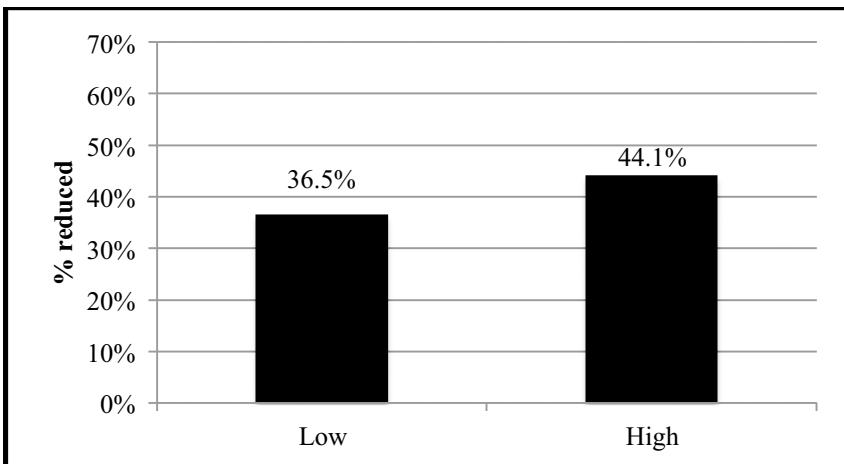
Figure 3. (%) Rates of *para* reduction by grammatical category of the following word
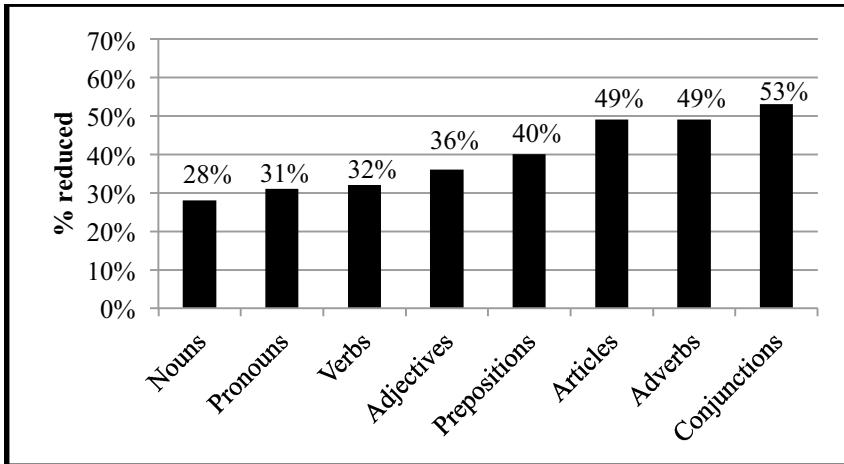
Figure 4. (%) Rates of *para* reduction according to the semantic meaning of the preposition
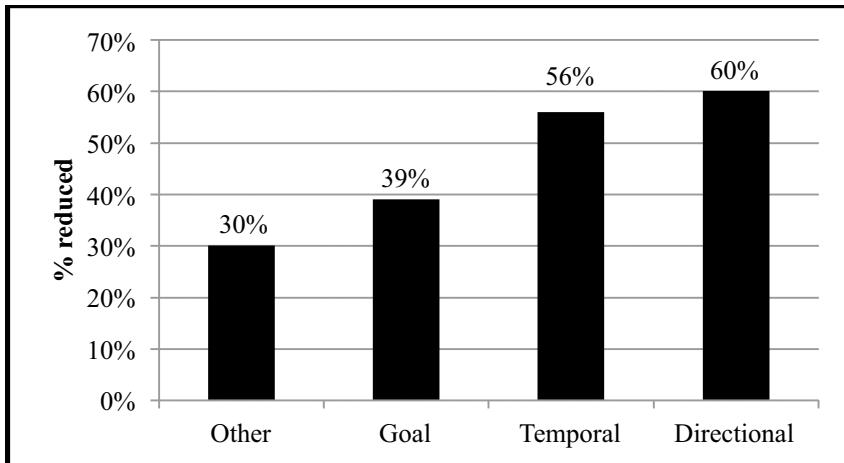

Figure 5. (%) Rates of *para* reduction by the frequency of *para*+WORD collocations

Moving on to the frequency factors, the high frequency *para*+WORD group (16-97 tokens) showed reduction in 44.1 percent (131/297) of the collocations and the low frequency group (1-15 tokens) showed 36.5 percent (182/498) reduction (Figure 5, above). Thus, there is an increase in reduction of *para* from the low frequency to high frequency groups for the *para*+WORD collocations.

For the WORD+*para* collocations (Figure 6, below), there was a 41.3 percent (100/242) rate of reduction in the high frequency group (4-25 tokens) and 38.5 percent (213/553) in the low frequency group (1-3 tokens), a difference that is smaller than for *para*+WORD collocations.

Figure 6. (%) Rates of *para* reduction by the frequency of WORD+*para* collocations
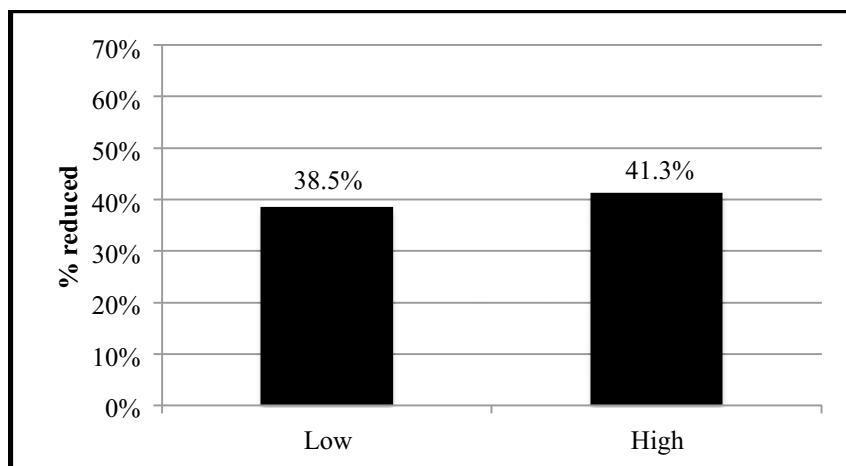

4.2. Social Factors

In order to compare the current study to previous variationist studies on Venezuelan Spanish, the social factors *age* and *sex* of the speaker were included. Figure 7, below, shows that for the *age* factor, the oldest speakers (aged 55 and above) favored reduction the most at 50.6 percent (180/356), followed by the youngest speakers (20-34) at 37.3 percent (95/255) and the middle-aged speakers (35-54) at 20.7 percent (38/184).

Finally, the results for sex of the speaker (Figure 8) show that in Málaga, women reduced *para* 41.4 percent (166/235) of the time while men reduced 37.3 percent (147/394) of the time.

Figure 7. (%) Rates of *para* reduction by age group.
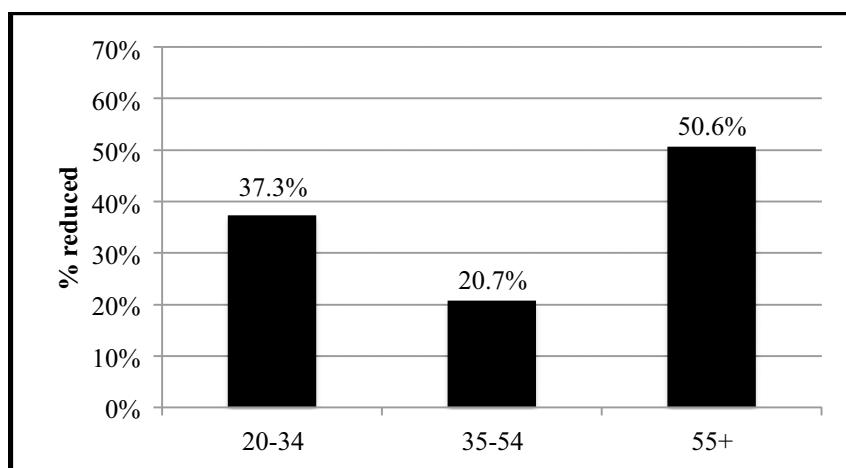
Figure 8. (%) Rates of *para* reduction by sex of the speaker
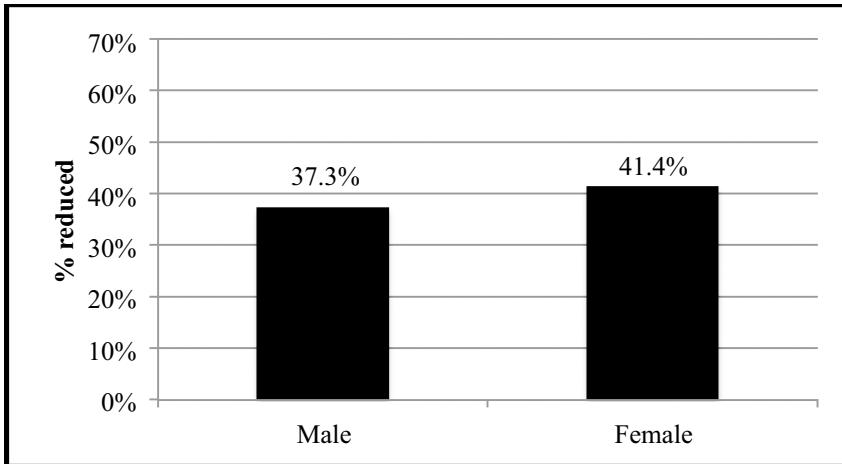
Before concluding with the results of the Varbul analysis using Goldvarb, it is necessary to resolve issues of interactions between factors and discuss why certain factors were excluded from the regression. As in Díaz-Campos et al. (2012), the phonological and grammatical category factors were found to interact with frequency and mask its effects. These factors, when included in the analysis, caused *para*+WORD frequency to be eliminated from the model. However, *para*+WORD frequency was included as a significant factor in the model when the phonological and grammatical category factors were excluded. Behind these interactions is the fact that six of the ten most frequent words following *para* in the corpus (Table 2, below) are a conjunction, an article, or an adverb, which are the three grammatical categories that coincide with the highest rates of *para* reduction. For example, the most frequent word following *para* in the corpus is the conjunction *que "that"* with 97 tokens (12.2 percent of the data, 50.5 percent reduction of *para*). As previously mentioned, this word is the primary reason why conjunctions, obstruents, and dorsals show very high rates of reduction[4]. Also, only six words (i.e., hacer, allá, acá, abajo, arriba, and atrás) make up 53.8 percent of back vowel and 65.2 percent of low vowel tokens, and also show the majority of *para* reduction in those categories. Thus, there is clearly an interaction between frequency and the phonological and grammatical category factors. Furthermore, there does not appear to be a true phonological effect on reduction because of no clear preference for consonants or vowels following *para* (40 percent reduction with a following vowel and 39 percent with a consonant). It is therefore very unlikely that conjunctions, obstruents, back vowels, dorsals or other phonological or grammatical categories are the primary force behind the pattern of variation. It is therefore argued here that collocation frequency is the linguistic variable that is driving the variation and not the phonological and grammatical category variables.

With these factors excluded, the results of the Varbul analysis (Table 3) show that the factors of preposition meaning, age, *para*+WORD frequency, and sex were selected as significant for the reduction of *para*. The factor WORD+*para* frequency is shown at the bottom in brackets because it was not selected as significant. The meaning of the preposition had the greatest effect (Range = 37). The directional meaning strongly favored reduction (.74), as well as the temporal meaning (.66). The goal meaning did not favor or disfavor reduction (.50) and all other meanings disfavored reduction (.37).

Age had the next strongest effect (Range = 34). The oldest age group favored reduction (.63), while the youngest age group did not favor or disfavor reduction (.49) and the middle age group strongly disfavored it (.29).

The *para*+WORD collocation frequency was the third factor (Range = 9). The high frequency collocations favored reduction (.56) and the low frequency collocations slightly disfavored it (.47).

---

[4] Dorsal was second only to palatal. Palatal segments only followed *para* five times in the data, indicating that the 60% reduction of *para* when a palatal followed is not a reliable indicator of their behavior.

Finally, sex of the speaker was selected as having an effect, albeit a smaller one (Range = 8). Women slightly favored reduction (.54) while men slightly disfavored it (.46).

Table 2. Ten most frequent words following *para* (in order), representing 37% of the data.

Word	Initial segment	Grammatical category
que 'that'	/k/	conjunction
el 'the' masc/sing	/e/	article
mí 'me'	/m/	pronoun
los 'the' masc/pl	/l/	article
hacer 'to do/make'	/a/	verb
qué 'what'	/k/	pronoun
la 'the' fem/sing	/l/	article
nada 'nothing'	/n/	adverb
ver 'to see'	/b/	verb
acá 'here'	/a/	adverb

Table 3. Results of the Varbrul analysis

Factors	Factor weight	% reduced	% of data	N
Meaning				
Directional	.74	59.7%	14.7%	117
Temporal	.66	55.6%	2.3%	18
Goal	.50	38.9%	48.8%	388
Other	.37	30.1%	34.2%	272
Range 37				
Age				
55 and above	.63	50.6%	44.8%	356
20-34	.49	37.3%	32.1%	255
35-54	.29	20.7%	23.1%	184
Range 34				
***para*+WORD frequency**				
High (16-97 tokens)	.56	44.1%	37.4%	297
Low (1-15 tokens)	.47	36.5%	62.6%	498
Range 9				
Sex				
Female	.54	41.4%	50.4%	401
Male	.46	37.3%	49.6%	394
Range 8				
WORD+*para* frequency				
High	[.52]	[41.3%]	[30.4%]	[242]
Low	[.49]	[38.5%]	[69.6%]	[553]

Log likelihood = -485.637 Significance = 0.049

4.3. A post-hoc analysis of the relationship between frequency and syntactic category

Given the findings of Jurafsky et al. (2001) and Shi et al. (2005) regarding frequency and syntactic category (content vs. function words), a post-hoc analysis was conducted to investigate possible effects of the interaction between frequency and the status of the word immediately following *para* as a content or function word.

There was an overall 44.3 percent (144/327) rate of *para* reduction for *para*+function collocations and 35.9 percent (168/472) in *para*+content collocations in the corpus. However, taking frequency and syntactic category into account together reveals an interesting picture. Specifically, high-frequency *para*+function collocations showed a 52.8 percent (103/195) rate of *para* reduction, whereas the high-frequency *para*+content collocations showed only 27.5 percent (28/102) reduction. Table 4 compares the rates of *para* reduction in specific collocations of *para* with high-frequency content and function words. What is evident is that the collocations of *para* with high-frequency function words consistently show over 43 percent reduction of *para*; whereas *para* with the four most frequent content words never exceeds 24.3 percent reduction.

Table 4. (%) Rates of reduction for high-frequency function and content words following *para*

Word following *para*	% *para* reduction	Function/Content status
que 'that'	50.5%	function
el 'the' masc/sing	43.5%	function
mí 'me'	24.3%	content
los 'the' masc/pl	51.9%	function
hacer 'to do/make'	24%	content
qué 'what'	47.6%	function
la 'the' fem/sing	47.6%	function
nada 'nothing'	15%	content
ver to see	21.1%	content

Further support for Jurafsky et al. (2001)'s findings comes from an examination of low frequency collocations. Low-frequency *para*+function collocations showed 31.8 percent (42/132) reduction and low-frequency *para*+content collocations showed 38.3 percent (140/366) reduction. Thus, the effect of function word status is not as clear at lower frequencies, which is what the previous literature has shown. Interestingly, the fact that low-frequency function words contribute less to *para* reduction than the low-frequency content words indicates that the higher rate of reduction in high-frequency function words compared to high-frequency content words is due primarily to their frequency of collocation with *para* rather than just their syntactic category. In other words, if syntactic category were the primary cause of higher rates of reduction of *para* when it is collocated with a function word compared to a content word, then there would be higher rates of reduction across the board for collocations of *para* with function words. But this is clearly not the case. Shi et al. (2005) found that different reduction processes respond in different ways to frequency and syntactic category. Specifically, t/d deletion responded more to syntactic category than frequency, a finding that goes against previous arguments by Bybee (2000) for the same phenomenon. However, they also found that palatalization processes in English were more responsive to word frequency than syntactic category. Also, vowel reduction was affected by both frequency and syntactic category. It is argued here, then, that the reduction of *para* is more affected by frequency than syntactic category, but that it is necessary to take both into account when discussing the reduction of *para*. Neither factor alone reveals the whole picture.

5. Discussion
5.1. Linguistic factors

The results of the current study show that multiple linguistic factors favor the reduction of the Spanish preposition *para* in Málaga, Spain. Interestingly, a statistically significant collocation frequency effect on the reduction of *para* was found for *para*+WORD collocations, as in Díaz-Campos et al. (2012). Specifically, high-frequency *para*+WORD collocations showed more reduction of *para* than low-frequency *para*+WORD collocations, supporting the claims of Bybee and Scheibman (1999) and Díaz-Campos et al. (2012) that high frequency collocations show more reduction than low frequency collocations. However, this effect was not significant for WORD+*para* collocations. The disparity between the size of the effect of *para*+WORD and WORD+*para* is similar to that found by Díaz-Campos et al. (2012). They explained it by the fact that WORD+*para* collocations are made up of a greater variety of words that each collocate with *para* less frequently, whereas the *para*+WORD collocations are made up of a smaller variety of words that each collocate with *para* more frequently. The current data show the same trend. It is argued here, as in Díaz-Campos et al. (2012), that the WORD+*para* collocations are therefore less able to obtain autonomy of storage in memory due to their lower frequency of collocation with *para* and have less opportunity for lexicalization and reduction of *para* compared to *para*+WORD collocations.

There is also another aspect of the frequency argument that had not been discussed prior to the current study in relation to *para*, namely the relationship between frequency and syntactic category. The current study supports the claims of Jurafsky et al. (2001) that probabilistic relations between words have a greater effect on function words than on content words. In other words, high-frequency collocations of *para* with function words (i.e., a function+function collocation) showed more reduction (52.8%) than high-frequency collocations of *para* with content words (i.e., function+content collocations, 27.5% reduction). Also, low-frequency *para*+function collocations showed less reduction of *para* than low-frequency *para*+content collocations. This supports previous claims that the effect of function word status is not as clear at lower frequencies, and it suggests that syntactic category alone cannot account for the reduction of *para.* This also supports Shi et al. (2005)'s claim that frequency and syntactic category affect phonological processes differently and should both be considered in an analysis. When syntactic category alone is considered, there is only an 8.4 percent difference between the reduction of *para* in collocations with content and function words (function>content) in the current study. However, when frequency is also taken into account, there is a much larger difference of 25.3 percent between reduction in collocations of *para* with function and content words (function>content) in the high frequency group. These results suggest that speakers store the frequency information regarding collocations of lexical items and that this, in tandem with a word's syntactic category status, contributes to the distribution of *para* reduction in the lexicon. However, it is argued here that frequency has a greater effect on the reduction of *para* than syntactic category. Similar effects have also been found in a study by Rao (2009) on the effects of frequency, repetition of words, grammatical category, word length, and position in the phonological phrase on intonational deaccentuation (a suprasegmental reduction process) in Barcelona, Spain. He found that more frequent words, words with fewer syllables, words recently repeated in discourse, and words occupying medial and initial positions in the phonological phrase were most likely to be deaccented (an F0 measure) in natural speech. Apart from his support of frequency effects, his results also support another finding by Jurafsky et al. (2001) that shorter words are reduced more than longer words. This coincides with Jurafsky et al. (2001)'s finding that function words tend to be shorter than content words. In these ways, Rao (2009) and the current study have founds similar results for segmental and suprasegmental reductive processes in Spanish.

While the results for frequency and syntactic category are interesting and there is certainly a difference in reduction for content and function words, especially in high frequency collocations, there is a potential problem that must be acknowledged and should be addressed in future studies. It should be noted that the high frequency function words following *para* are almost all definite articles. This study did not take into account the noun that follows the definite article as part of the collocation with *para*+*definite article*. It makes sense according to Exemplar Theory (Pierrehumbert, 2001) that

para+verb would make a collocation in memory, but seems strange that *para+definite article* would be stored without a noun. Because of this, the frequency effects evidenced in this study for *para+article* may actually reflect *para+article+noun*. Consequently, future studies should include the noun in the coding of these collocations.

The final linguistic factor that was significant was the meaning of the preposition[5]. As in Díaz-Campos et al. (2012), Bentivoglio et al. (2005) and Guirado (2007), a directional meaning of the preposition *para* highly favored the reduction of *para*. However, the current study departs from previous studies in that it shows that a temporal meaning was the second-most favoring of reduction. It is interesting, though, that in all four of these studies, the temporal category has consisted of very few tokens relative to the other categories of prepositional meaning. Therefore, it is doubtful that any of the studies examining this factor have been able to provide a truly reliable indicator of the reduction of *para* as a temporal preposition.

5.2. Social factors

The social factors selected by the Varbrul analysis were mostly consistent with previous research. The age of the speaker was significant and the result was partially consistent with previous sociolinguistic studies by Bentivoglio et al. (2005) and Guirado (2007), which also showed that older speakers reduced *para* the most. However, the younger speakers reduced more than the middle-aged speakers in the current study. The u-shape curve could be evidence of age grading, that the middle-aged speakers, because of their position in life (e.g., employments status), may be using more standard variants than the older and younger speakers. However, a study more dedicated to social factors would be necessary to verify this interpretation. Given that the age factor had the second-largest effect in the current study, and that socioeconomic status had the strongest effect in Bentivoglio et al. (2005) and Guirado (2007), it appears that social factors in general exert a great influence on the reduction of *para* cross-dialectally.

There also seems to be a very slight difference between men and women in the three studies to date that have included this factor, with women reducing more than men. In Guirado (2007), women reduced ten percent more than men, and in Bentivoglio et al. (2005), women reduced only two percent more than men. In the current study, women reduced 4.1 percent more than men. Still, the women are not acting in a very innovative manner overall. Their reduction of *para* (41.4%) is far from categorical and is not different enough from the men (37.3%) to be able to suggest social implications. To know more conclusively why women are reducing as much as and slightly more than men, a more detailed study with different social classes and speech styles would be necessary.

6. Conclusion

The current study has investigated contributing factors to the phonetic reduction of the Spanish preposition *para* in Málaga, Spain. This is the first study to investigate the effects of frequency, syntactic category, and sociolinguistic factors together on the reduction of Spanish *para*, and also the first variationist study of Spanish *para* reduction outside of Venezuela. An analysis of 795 tokens of *para* from the PRESEEA corpus of Málaga speech (Villena Ponsoda et al., 2002) revealed that *para* in Málaga is variable and its reduction is conditioned by various factors.

---

[5] An important question is whether there is any co-linearity (i.e., overlap) between the meaning of the preposition and *para+WORD* frequency, as was found for the phonological and grammatical category factors. For example, *para hacer* was a high frequency collocation and also represented a goal meaning of *para*. However, examination of the data does not indicate that co-linearity is a problem with these two factors. The ten most frequent collocations represent various meanings of *para*. *Para que, para hacer, para qué,* and *para ver* represent a goal meaning of *para*, but *para el, para los,* and *para la* represent multiple different meanings depending what follows the article, but they never represent a goal meaning. Also, *para mí* and *para nada* represented the 'other' category, and *para acá* carries a directional meaning. In sum, the ten most frequent collocations with *para* represent a variety of prepositional meanings, and there does not appear to be an imbalance toward one prepositional meaning for the highest frequency items. Furthermore, there did not appear to be an interaction between these factors in the Varbul analysis.

It was shown that when *para* collocates frequently with a following word, *para* is reduced more often than in low-frequency collocations. It was also shown that including both frequency and syntactic category of the word following *para* shows a clearer picture of frequency effects. When p*ara* collocates frequently with a function word it reduces 25.3 percent more than when collocated with a highly frequent content word, and also reduces more than in low frequency collocations with content and function words. The author agrees with Bybee and Scheibman (1999) and Jurafsky et al. (2001) that speakers are aware at least subconsciously of the probabilistic relations between words and the frequency with which words collocate, and that the frequency of a collocation and the syntactic category of the words in the collocation contribute together to the likelihood that *para* will be reduced.

Also, when *para* carries a directional meaning (e.g., "Vamos pa(ra) Málaga" "*We are going to Málaga"),* reduction is cross-dialectally most likely to occur compared to other meanings. Other significant predictors of reduction are age (55 years or more) and the sex of the speaker (female>male). Taking into account the research to date on this phenomenon in Spanish, it can be concluded that the most likely individuals to reduce *para* are the lower socioeconomic classes (although higher class speakers in Málaga also reduce), older generations (55+) and women (albeit slightly more than men).

This paper contributes to our understanding of frequency effects on phonological processes, and extends the previous discussion of the phonetic reduction of *para* to include the effects of syntactic category and its relationship to frequency effects. This study also broadens the scope of previous studies to include a dialect outside of Venezuela, and included social factors for comparison to Venezuela. Many similarities were found between the studies to date, providing more evidence for the claims made thus far.

Future research on the effects of frequency on Spanish phonetic reduction processes should take syntactic category (i.e., function/content status) into account alongside frequency factors for a more complete description of the phenomenon. Furthermore, larger corpora are ideal to be able to test the gradient nature of frequency effects as Díaz-Campos et al. (2012) did. Likewise, an analysis using a program that allows for continuous dependent variables, such as Rbrul, could eliminate the need for making arbitrary frequency divisions and give a better understanding of the gradient nature of frequency effects. Other statistical methods could also shed light on important interactions between independent variables that are not able to be sufficiently investigated using Goldvarb. Future studies should also take into account the position of each token in the phonological phrase or sentence, given the findings of Rao (2009) (see also Fougeron and Keating, 1997) regarding the influence of phrase position on reductive processes. Finally, a larger-scale analysis of the PRESEEA Málaga corpus including all of the interview data, if available, would be helpful in interpreting the effects of social factors, and would allow the inclusion of more socioeconomic classes.

References

Bentivoglio, Paola, Guirado, Krístel, & Suarez, Guillermina. (2005). La variación entre *para* y *pa* en el habla de Caracas. Homenaje a Montes Giraldo. *Estudios de Dialectología, Lexicografía, Lingüística general, Etnolingüística e Historia Cultural*. Bogotá: Instituto Caro Cuervo. 2005, 214-237.

Bentivoglio, Paola, & Sedano Mercedes. (1993). Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana. *Boletín de Lingüística* 8. 3-35.

Bybee, Joan. (2002a). Phonological Evidence for Exemplar Storage of Multiword Sequences. *Studies in Second Language Acquisition, 24*(2), 215-221.

Bybee, Joan. (2002b). Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change. *Language Variation and Change, 14*(3), 261-290.

Bybee, Joan. (2006). From usage to grammar: the mind's response to repetition. *Language* 82 (4). 711-733.

Bybee, Joan, & Scheibman, Joanne. (1999). The Effect of Usage on Degrees of Constituency: The Reduction of Don't in English. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 37*(4), 575-596.

Calcaño, Julio. (1949). *El castellano en Venezuela*. Caracas: Ministerio de Educación.

Chela-Flores, Berta, & Gelman, Jeannette. (1988). *El habla de Maracaibo*. Maracaibo: Universidad de Zulia.

Cuervo, Rufino. (1939). Apuntaciones críticas sobre el lenguaje bogotano. Bogotá: Editorial El Gráfico.

Cuervo, Rufino. (1944). Castellano popular y castellano literario. Bogotá: Instituto Caro y Cuervo.

Díaz-Campos, Manuel, Fafulas, Stephen, & Gradoville, Michael. (2011). Going retro. An analysis of the interplay between socioeconomic class and age in Caracas Spanish. In *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 65-78).

Díaz-Campos, Manuel, Fafulas, Stephen, & Gradoville, Michael. (2012). Variable degrees of constituency: Frequency effects in the alternation of pa vs. para in spoken discourse. In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, ed. by Kimberly Geeslin and Manuel Díaz-Campos. Somerville, MA: Cascadilla Proceedings Project.

Díaz-Campos, Manuel, & Gradoville, Michael. (2011). An analysis of frequency as a factor contributing to the diffusion of variable phenomena: Evidence from Spanish data. In *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, ed. by Luis Ortíz López. Somerville, MA: Cascadilla Proceedings Project.

Díaz-Campos, Manuel, & Ruiz-Sanchez, Carmen. (2008). The Value of Frequency as a Linguistic Factor: The Case of Two Dialectal Regions in the Spanish Speaking World. In *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*, ed. by Maurice Westmoreland and Juan Antonio Thomas, 43-53. Somerville, MA: Cascadilla Proceedings Project.

Dominguez, Carmen Luisa. (1996). El habla de Mérida: Un corpus de estudio. *Lengua y Habla* 1,2.

File-Muriel, Richard. (2007). *The role of lexical frequency and phonetic context in the weakening of syllable-final lexical /s/ in the Spanish of Barranquilla, Colombia*. Indiana University PhD Dissertation.

File-Muriel, Richard. (2009). The role of lexical frequency in the weakening of syllable-final lexical /s/ in the Spanish of Barranquilla, Colombia. *Hispania*, 348-360.

Fougeron, Cécile, & Keating, Patricia. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*, 3728.

Gradoville, Michael. (October, 2012). *Collocation frequency and the reduction of Carioca Portuguese para.* Paper presented at the 41st *New Ways of Analyzing Variation*, Bloomington, Indiana.

Henríquez Ureña, Pedro. (1940). *El español en Santo Domingo*. Buenos Aires: Universidad de Buenos Aires.

Hooper, Joan. (1976). Word frequency in lexical diffusion and the source of morphophonological change. *Current progress in historical linguistics*, ed. by William M. Christie, 96-105. Amsterdam: North Holland.

Jurafsky, Daniel, Bell, Alan, Gregory, Michelle, & Raymond, William. (2001). Probabilistic relations between words. Evidence from reduction in lexical production. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 229-54. Amsterdam: John Benjamins.

Kenstowicz, Michael. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Rao, Rajiv. (2009). Deaccenting in spontaneous speech in Barcelona Spanish. *Studies in Hispanic and Lusophone Linguistics*, *2*(1), 31-75.

Samper Padilla, José Antonio. (2011). Sociophonological variation and change in Spain. *The Handbook of Hispanic Sociolinguistics*, 98-120.

Sankoff, David, Tagliamonte, S., & Smith, E. (2005). *Goldvarb X*. Toronto: Department of Linguistics, University of Toronto. Online: http://individual.utoronto.ca/tagliamonte/Goldvarb/GV index.htm.

Shi, Rushen, Gick, Bryan, Kanwischer, Dara, & Wilson, Ian. (2005). Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research*, *34*(4), 341-364.

Vida, Matilde. (2005). Estudio sociofonológico del español hablado en la ciudad de Málaga. Condicionamientos sobre la variación de /-s/ en la distensión silábica. Alicante: Universidad.

Villena Ponsoda, Juan Andres; Ávila Muñoz, Antonio; Cuevas Molina, Inmaculada; Díaz Montesinos, Francisco; and Vida, Matilde. (2002). PRESEEA - Málaga, http://preseea.linguas.net/Equipos/M%C3%A1laga.aspx.

Selected Proceedings of the
16th Hispanic Linguistics Symposium

edited by Jennifer Cabrelli Amaro, Gillian Lord, Ana de Prada Pérez, and Jessi Elana Aaron

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Bedinghaus, Rob. 2013. ¿Vas pa(ra) Málaga? The Reduction of *para* in Málaga, Spain: Effects of Frequency, Syntactic Category, and Social Factors. In *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, ed. Jennifer Cabrelli Amaro et al., 238-252. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2937.