Variable Degrees of Constituency: Frequency Effects in the Alternation of *pa* vs. *para* in Spoken Discourse

Manuel Díaz-Campos, Stephen Fafulas, and Michael Gradoville
Indiana University

1. Introduction

Bybee and Scheibman (1999) make a distinction between "frequently used structures" and "regular structures." Their division is rooted in the hypothesis that frequently used structures gain autonomous storage status, thus rendering the individual analysis of each element of the structure unnecessary for its processing by speakers. In essence, these frequently used structures become something like "fused-units" in the lexicon, most likely due to a cognitive drive for more efficient processing and memory, facilitating both their storage and retrieval in discourse. Use, repetition, and frequency determine the degree of constituency of these units and, for this reason, frequent combinations are posited to have a tighter constituent structure than less frequent combinations. Bybee and Scheibman (1999) emphasize their point by showing that *don't* in American English undergoes more reduction in very frequent combinations, such as *I don't know*, an instance of a structure which has acquired autonomy of storage, making it different from less frequent contexts in which *don't* is used like *we don't know*.

In this study we test similar constituency effects in Venezuelan Spanish. Recent work analyzing the variation between *para* and *pa* 'for' in Spanish by Bentivoglio, Guirado and Suárez (2005) and Guirado (2007) indicate that this is a productive alternation in all areas of Venezuela observed, namely Caracas, Maracaibo, and Mérida. These results reveal that reduction to *pa* is more likely to occur when the preposition has a directionality meaning associated with it, when it is followed by a consonant, and when the speaker is 60 or more years old and from a lower socioeconomic background. The present analysis builds on these findings by incorporating an analysis of frequency of use. In our study we specifically examine how repetition in discourse determines variable degrees of constituency between *para* and surrounding words, thus predicting the reduction of *para* in cases where the degree of constituency is tightest. Our study contributes to the overall understanding of the role of usage in explaining reductive phonetic processes in contemporary Spanish.

2. Previous Studies
2.1. Alternation of the preposition *para*

While mention of the reduction of the preposition *para* 'for, in order to' (e.g. *Mañana me voy para ~ pa Miami* 'Tomorrow I leave for Miami') appears in Spanish dialectal studies (e.g. Cuervo 1939, 1944; Calcaño 1949; Henríquez Ureña 1940), little empirical investigation of this variable exists. Notable exceptions are Cedergren (1973) and Garcia (1979). Cedergren (1973) described the phenomenon as deletion of intervocalic /ɾ/ and observed that deletion is more common in lower socioeconomic groups, informal style, the youngest group in the sample and female speakers. Garcia (1979) tested the perception of *para* usage between two Spanish-speaking groups: Chicano speakers living in El Paso, Texas and Mexican speakers living in Juárez, Chihuahua**.** The author introduces the variable by stating that "the reduction of *para* to *pa* is quite common all over the Spanish-speaking world" and the two forms "are commonly thought to be in free variation" (106). While the results were preliminary, they revealed that the reduced form *pa* was acceptable to both groups in a large number of contexts, although to differing degrees.

To our knowledge, the first study to systematically analyze the alternation between the full form *para* and the reduced form *pa* is Bentivoglio et al. (2005) and Guirado (2007). Bentivoglio's et al. data come from 48 sociolinguistic interviews of the *Corpus sociolingüístico de Caracas 1987* (Bentivoglio & Sedano 1993). They note that intervocalic /ɾ/ in *para* tends to reduce in many Spanish dialects (especially in constructions like *para allá* > [payá], see also Willis and Bradley 2008). The socio-cultural observation lies in the assumption that while both forms exist in the speech of all socioeconomic classes it is more heavily associated with informal speech.

For their analysis the participants were equally divided by age (30-45, 60+), socioeconomic level (upper, middle, lower), and sex (male, female). The linguistic factors they included in their study were: following phonological context (vowel vs. consonant) and preposition meaning (purpose, directionality, temporality). In total they coded 1599 tokens, rendering an overall distribution of 787 (49%) for the full form (*para*) and 812 (51%) for the reduced form (*pa*). Particularly important to the current study is the fact that the authors excluded forms such as *pa(ra) + aquí, allá, abajo*, etc. assuming that these constructions were invariant. The results of the multivariate analysis conducted with GoldVarb showed that lower socioeconomic level speakers favored the reduced form (.80), as did the 60+ age group (.54). Further, the factors following consonant (.54) and directionality (.66) favored the reduced form *pa*.

Guirado (2007) further analyzed the *para-pa* alternation. In her study the author coded for the same linguistic and extralinguistic factors as Bentivoglio et al. (2005), but this time with the additional factor of dialect. The 72 speakers included in the study were from three distinct regions of Venezuela: 24 from the *Corpus sociolingüístico de Caracas*, 24 from the *Corpus sociolingüístico de Mérida*, and 24 from the *Corpus sociolingüístico de Maracaibo*. In total her analysis consisted of 2144 tokens, 1026 (48%) of the reduced form *pa* and 1118 (52%) of the full form *para*. The patterning of her results was almost identical to that found by Bentivoglio et al. (2005). The factor groups selected by GoldVarb were (factors favoring reduction to *pa* are in parentheses): socioeconomic class (lower = .80), semantics of the phrase associated with *para* (directionality = .75; temporality = .55), following phonological segment (consonant = .54), age (60+ = .54), and, lastly, dialect (Merida = .54 and Caracas = .50).

2.2. Degrees of constituency and frequency effects

A number of studies have pointed to a persistent pattern of language in that phonetic reduction advances more rapidly in more frequently used words/constructions (Bybee & Scheibman 1999; Bybee 2001, 2010; Hooper 1976; Jurafsky, Bell, Gregory, and Raymond 2001). Bybee (2002) shows that English final *t/d* deletion, as well as Spanish [ð] deletion, is highly sensitive to frequency effects. Bybee holds that there are physiological motivations for these patterns: "[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" (2006:715). This is what Bybee (2006) refers to as the *reducing effect* of frequency. Following this claim, several authors have found evidence for the model in Spanish: syllable-final /ɾ/ reduction in Venezuelan and European Spanish (Díaz-Campos 2005, 2006; Ruiz-Sánchez 2006; Díaz-Campos & Ruíz-Sánchez 2008); vowel coalescence in New Mexican Spanish (Alba 2006); intervocalic /d/ reduction in New Mexican Spanish and Venezuelan Spanish (Bybee 2002; Díaz-Campos & Gradoville 2011); and syllable-final /s/ reduction in several varieties of Spanish (Brown 2008; File-Muriel 2007).

As mentioned in the introduction, Bybee & Scheibman (1999) found that higher frequency constructions, such as *I don't know*, evidenced the highest rate of *don't* reduction in their data. They posit that speakers do not independently access *I*, *do*, *not*, and *know* but rather that this combination is in fact stored and accessed as a whole unit. As fluency increases then, lexical strings become processed as single units (Boyland 1996). This is what Bybee (2006) refers to as *autonomy* of storage. The reducing effect can also be applied to these strings of words. Every time a fixed routine or sequence of words is repeated the articulatory gestures overlap and a resultant reduction occurs.

Extending beyond the level of phonology alone, Bybee's (2001, 2006) usage-based model of language holds that our mental representations of linguistic categories are heavily dependent on our

linguistic experience. Crucially, our cognitive representations are altered based on the frequency with which individual units of language are used, as well as the specific contexts in which they appear.

Returning again to phonetic reduction, much of Bybee's (2001) model is built on the articulatory gesture-based account of reductive processes argued for by Pagliuca and Mowrey (1987), who divide reductive processes into two categories: substantive and temporal. *Substantive reduction*, involves a decrease in the magnitude of an articulatory gesture. *Temporal reduction*, involves an increase in the overlap of articulatory gestures. We hold that Spanish intervocalic /ɾ/ reduction in the word *para* falls into the substantive category, given that the change can be viewed as the gradual decrease in the magnitude of the tongue's gesture[1]. Further, we argue that our data can best be interpreted in light of Bybee's *reducing* and *autonomy* effects. That is, we hypothesize that intervocalic /ɾ/ reduction in the word *para* will be greater in those instances of frequently-used combinations occurring in similar linguistic environments.

We add to this previous work on the phonetic reduction of highly frequently used words/constructions (Bybee 2001, Jurafsky et al. 2001) by analyzing the reduction of *para* in Venezuelan Spanish in light of frequency of use and surrounding linguistic environment. In this study we set out to address three specific research questions:

1. Does the phonetic context following *para* affect its reduction?
2. What is the effect of semantic meaning on *para* reduction?
3. What is the effect of frequency of co-occurrence with items immediately preceding and following *para*?

3. Methodology

The data of the current study come from the corpus *Estudio sociolingüístico de Caracas* (Bentivoglio and Sedano 1993). In its entirety the corpus is comprised of 160 semi-directed Labovian-style sociolinguistic interviews, all of which were conducted between 1987 and 1988. All participants included in the corpus were born and raised in Caracas with the additional inclusion criterion that their parents also be from Caracas. The speakers are evenly stratified by age (14 – 29, 30 – 45, 46 – 60, 61 and over), gender (male, female), and socioeconomic group (upper, upper-middle, middle, lower-middle, and lower). We used data from all participants in the corpus.

For the analysis we exhaustively extracted all tokens of the dependent variable, full and reduced forms of *para*, and coded each token for the following independent linguistic variables:

1. Following phonetic context: this variable was included following previous investigations (e.g. Bentivoglio et al. 2005, Guirado 2007, etc), but we decided to examine more detailed categories by considering a following vowel height, frontness/backness; or if a consonant, we consider its manner and place of articulation. In example (a) *pa* is followed by the consonant [l] which was coded as *alveolar*, *lateral*.

a. Playa también, más que todo siempre íbamos *pa' la* playa (speaker -cb3mb_87)
   'Beach also, more than anything else we always went to the beach.'

2. Meaning associated with the preposition (i.e. goal, directionality, and temporality): Bentivoglio et al. (2005) proposed a classification of the meaning associated with the context where the preposition is used following De Bruyne (1999). Specifically, Bentivoglio et al. (2005: 227) categorized the cases as: 1. Goal, 2. Direction, and 3. Temporal. We also included a fourth category "other" for ambiguous cases difficult to label.

---

[1] Although it is likely that there is a corresponding temporal reduction of the formerly two /a/ sounds once the substantive reduction has occurred. However, proving this point would require durational measurements.

a. Goal
Para comprarse un par de patín winchester. (Speaker- cc4ma_87)
'To buy himself a pair of Winchester skates.'

b. Direction
Un campesino se viene *pa'* Caracas. (Speaker- cb3mb_87)
'A farmer comes to Caracas.'

c. Temporal
En esa empresa trabajé diecinueve años. diecinueve años con ... on.., once meses *para* veinte.
(Speaker- cd5ma_87).
'In that company, I worked 19 years with eleven months, going on 20.'

d. Other
Porque tú sabes que los míos están grandes y ya es un problema que viene es *para* ti.
(Speaker- cb5fa_87)
'Because you know mine are grown, and the problem that is coming is for you.'

3. Frequency of the chunk: based on the work of Bybee and Schiebman (1999: 577), we decided to examine how *para* combines with other constituents following the idea that "the more often two elements are used together, the more tightly they will be fused or bonded phonologically and semantically." Frequency of the chunk was determined by first tabulating the number of occurrences of each *para* + WORD combination and WORD + *para* combination. For each frequency of co-occurrence variable five categories were initially established, each of which included approximately 20% of the tokens. Categories were subsequently collapsed to resolve interactions.

3.1 Frequency of co-occurrence of *para* + WORD. The three most frequent collocations

a. Pero como él insistía ir *pa' la* casa. (Speaker- cb5fd_87)
   'But since he insisted to go to the house.'

b. Vamos *pa' el* amazonas. (Speaker- ca2mc_87)
    'We're going to the Amazon.'

c. *Para que* aquello funcione.(Speaker- cd4mb_87)
   'So that it works.'

3.2 Frequency of co-occurrence of WORD + *para*. The three most frequent collocations

a. Me *iba para* la técnica. (Speaker- cb4mb_87)
   'I went to technical school.'

b. Pal liceo *y pal* colegio. (Speaker- ca5fb_87)
   'To school and to private school.'

c. Todavía no he conocido esa *zona pa'* ir pal parque. (Speaker- cb4fd_87)
   'I don't know that zone yet to go to the park.'

4. Following grammatical category (i.e. adverb, determiner, noun, preposition, etc.): the grammatical category of the word following *para* was taken into account to determine the most common patterns of distribution found in the data.

a. Pronoun
Para tú contar con una tierrita... (Speaker- cb2fd_87)
'For you to have a small land...'

b. Infinitive
Para ir a comer mango... (Speaker- cc5ma_87)
'To go to eat mangoes...'

c. Adjective
Pa mi casa porque de momento... (Speaker- ca5mc_87)
'To my house because at the moment...'

d. Noun
Fui *pa Caracas*... (Speaker- cc5fd_87)
'I went to Caracas...'

e. Determiner
Vamos *pa la* playa... (Speaker- ca1mc_87)
'We go to the beach'

f. Conjunction
Pa que mi mamá no supiera... (Speaker- cc2fb_87)
'so that my mom would not find out'

g. Preposition
Pero *pa hacia* el lado mío... (Speaker- cd5fa_87)
'But toward my side...'

h. Adverb
Y *para arriba* ¿no?.. (Speaker- cb1md_87)
'And above, no?'

4. Results
4.1. Distributive frequencies

In total we analyzed 5,310 tokens of *para* for phonological context, frequency, and semantic meaning. Table 1 displays the overall distribution of the full and reduced forms in the data. Of the total 5,310 tokens, reduction represents 38.5% of the data, while retention of the full form represents 61.5%.

Table 1. Distribution of variants of *para* in a corpus of Caracas Speech based on 160 participants

	Reduced	**Retention**	**Total**
Tokens	2044	3266	5310
Percentage	38.5%	61.5%	100%

Next, we turn to the rate of use of the reduced variant based on phonological context. As can be seen in Figure 1, *para* has a stronger tendency to reduce in our data when the following segment is a low vowel or obstruent. This tendency does not seem to correlate to any phonological pattern, since the most favoring categories are on the opposite ends of typical scales used to classify segment sonority within traditional phonological approaches (Kenstowicz 1994). In what follows, we present further evidence that these trends are not indicative of a particular phonological pattern, but rather a result of frequency of use.

Figure 1. (%) Rates of reduction of *para* in Caracas Spanish by manner of articulation of following segment
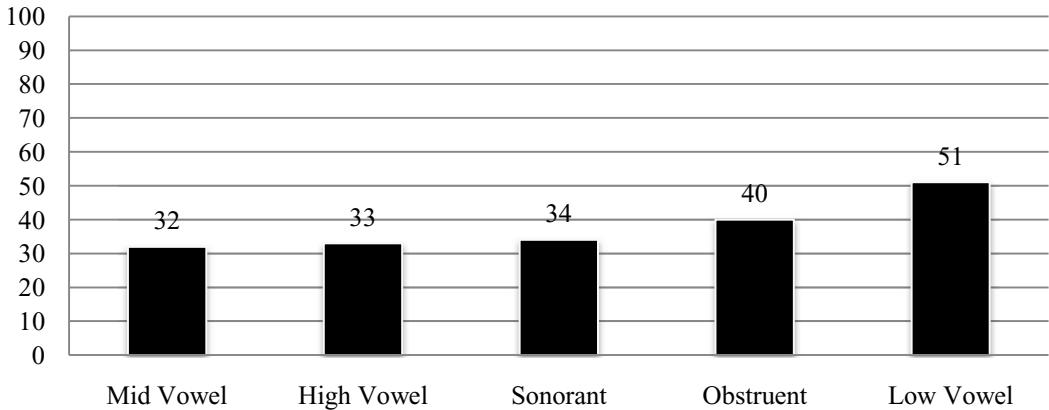
Regarding the place of articulation of the following segment (Figure 2), *para* reduces more frequently when the following segment is a back vowel or dorsal. Again, we present evidence that this apparent trend is not really phonologically driven, as it overlaps with the results of frequency discussed below.

Figure 2. (%) Rates of reduction of *para* in Caracas Spanish by place of articulation of following segment
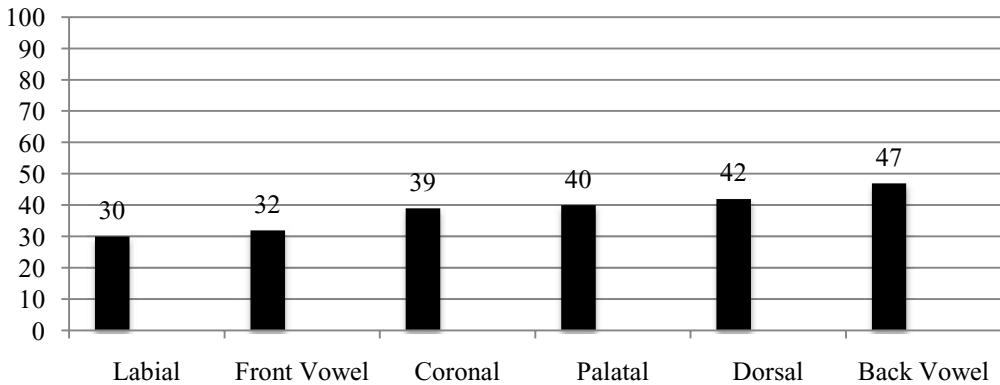

Figure 3 displays the rates for reduction of *para* based on the grammatical category of the following word. We can observe that the highest rate of reduction occurs when *para* occurs before adverbs, prepositions, conjunctions, and determiners.

Figure 3. (%) Rates of reduction of *para* in Caracas Spanish by grammatical category of following word.
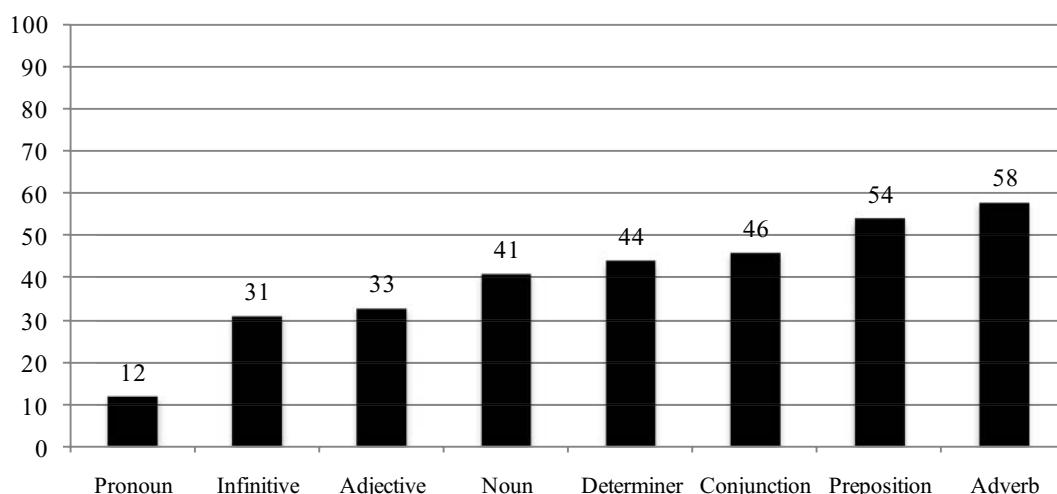

Figure 4 shows the patterns of reduction according to preposition meaning. Reduction of *para* is strongly favored when the preposition functions to indicate directionality. This is largely in line with the results of Bentivoglio et al. (2005) and Guirado (2007) who found a similar effect for reduction based on meaning of the preposition in their study.

Figure 4. (%) Rates of reduction of *para* in Caracas Spanish by meaning of preposition
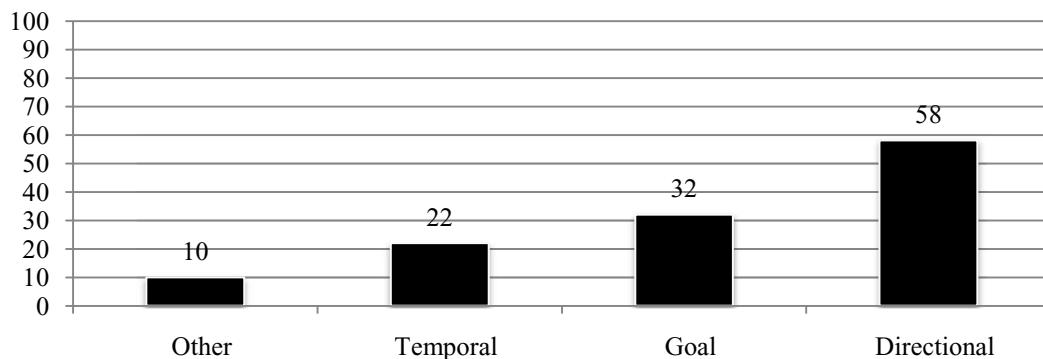
We now turn our attention to the variables that have not before been accounted for in predicting the reduction of *para,* those of frequency of use. Observe the results in Figure 5. What becomes immediately apparent is that the reduction of *para* increases in tandem with the frequency of co-occurrence of the following word. We return to further explore this finding in the discussion section below.

Lastly, we draw attention to the observed effects for reduction of *para* based on the frequency of co-occurrence of the previous word. Again, although in somewhat of a less dramatic fashion, Figure 6 provides preliminary evidence of a frequency effect in our data for the WORD + *para* collocation, as was also evidenced for the *para* + WORD collocation. That is, reduction of *para* increases in direct correlation with the frequency of co-occurrence of the words immediately preceding and following the preposition.

Figure 5. (%) Rates of reduction of *para* in Caracas by frequency of *para* + WORD collocation[2]
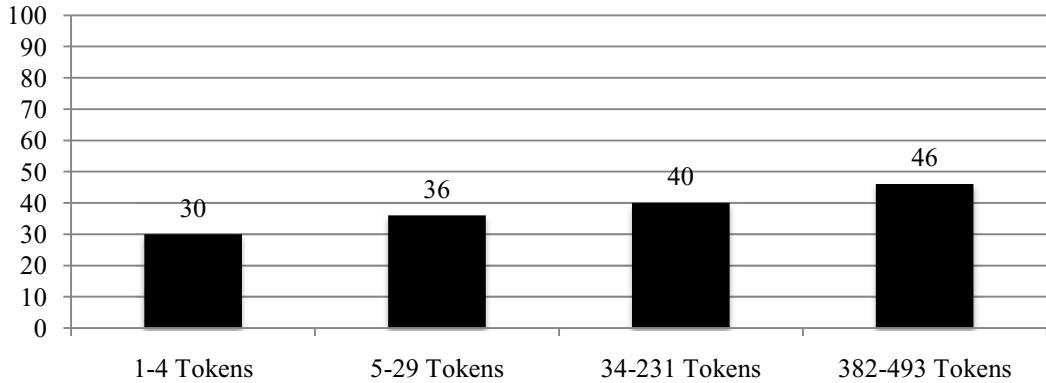
Figure 6. (%) Rates of reduction of *para* in Caracas by frequency of WORD + *para* collocation
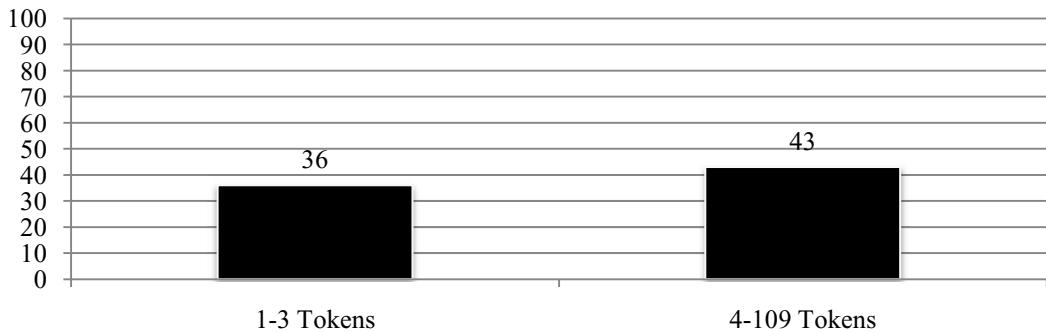
4.2. Degrees of constituency and frequency effects for para

To wrap up the results section, we will present the findings from the GoldVarb binomial regression. However, certain issues must be resolved beforehand regarding the interactions apparent between certain variables. First, regarding the apparent phonological patterning, we must keep in mind that the most frequent words are phonologically similar. Observing Table 2 we can see that the nine words most-frequently following *para* account for merely five sounds: /k/, /l/, /e/, /m/, and /a/. These nine words represent approximately two-fifths of the tokens in the dataset, meaning that these words have a disproportionate effect on the trends observed in Figures 1 and 2.

---

[2] A gap exists between categories of *para* + WORD because there are no collocations with such frequencies in the gap; for example, there are no collocations of frequency 232-381 or 30-33. So, 382 represents the frequency of the least frequent collocation of the most frequent category and 231 represents the most frequent collocation of the second most frequent category.

Table 2. Nine most frequently co-occurring words following *para* in the corpus, representing 40% of the data.

Word	First Sound	Function
que 'what/that'	/k/	conjunction
la 'feminine article'	/l/	determiner
el 'masculine article'	/e/	determiner
mí 'me'	/m/	pronoun
acá 'here'	/a/	adverb
allá 'there'	/a/	adverb
los 'them'	/l/	determiner
hacer 'to do/make'	/a/	verb
arriba 'above'	/a/	adverb

Table 3 shows the rates of reduction of *para* for all tokens before a word that starts with /k/, /l/, /e/, /m/, and /a/, and also the reduction rates when only 40% of the tokens are included, representing the least frequent co-occurring words in the dataset. What is immediately evident is that the rates of reduction are dramatically lessened when we remove the frequently co-occurring words. That is, if we observe those instances of the *para* + WORD collocation in which the initial segment of the following word is an /a/ the overall rates of reduction are 51.1%, while when we remove the instances of frequent words from the collocation, a notable difference of 10.7% of reduction is unveiled. Thus, the most frequently occurring words of the *para* + WORD collocation with following /a/ (especially *arriba, allá, hacer, acá*) account for a disproportionately high share of reduction where the following word of the *para* + WORD collocation begins with /a/[3]. Further, if we consider the two most frequent words with an initial segment /l/ an even more striking difference of 32.8% is observed. This trend holds for all but one of the following initial segments in the *para* + WORD collocation displayed in Table 3. The fact that the initial segment /m/ in *mí* contradicts the pattern observed to this point is not strong enough, in and of itself, to negate the hypothesis. It does, however, warrant our attention. Although this will need to be analyzed in greater detail in a future study, a possible explanation is that *mí* is a stressed pronoun mostly found in clause final position. Thus, it is plausible that the contradiction of the trend has to do with a phrasal boundary effect, given this particular collocation is more prone to appear in a position which is less likely to reduce. Stemberger and Bernhardt (1999) highlight that certain positions are of high activation, a finding which is backed by psycholinguistic research. We argue that a phrasal boundary may be regarded as a high activation domain given its salient position. However, this will need to be tested in a future study.

Table 3. Rates of reduction for *para* in the corpus by following phonological segment

Sound	All Tokens	With Most Freq. Removed	Diff
/k/	42.7%	33.3%	-9.4%
/l/	46.1%	13.3%	-32.8%
/e/	32.1%	21.3%	-10.8%
/m/	20.2%	33.6%	+13.4%
/a/	51.1%	40.4%	-10.7%

Returning our attention to Table 2 where the nine most frequent words following *para* are displayed, we find evidence for a strong correlation between frequency of use and grammatical category, since an overwhelming majority of these nine words are adverbs and determiners. However, we hold that frequency of the collocation explains the results more satisfactorily than the variables of phonetic context or grammatical category. That is, the effects of following grammatical category and phonological segment are epiphenomenal in nature and the true driving force behind the reduction of

---

[3] A similar procedure excluding all uses of *para* with directional meaning results in a drastic decrease in the reduction rate of *para* when followed by /a/, meaning that the higher rates of reduction of the low and back vowels in Figures 1 and 2 should be attributed to this interaction, as well as that with frequency of use.

para is the frequency with which it co-occurs with a following word/segment. As Bybee explains, "the principle experience that triggers chunking is repetition (…) one effect of chunking in production is the overlap and reduction of articulatory gestures" (2010: 34).

We would like to call the reader's attention to the fact that, given the lack of effect for phonetic context. We excluded the variable phonetic context since in our preliminary runs we determine that this variable statistically interacts with frequency of use. Further, we also excluded following grammatical category, since it confounds both preposition meaning and frequency of use.

Table 4 reveals that the factor groups of (a) preposition meaning, (b) *para* + WORD frequency, and (c) WORD + *para* frequency were selected as significant in predicting the reduction of *para*. The range of 55 indicates that preposition meaning has the greatest magnitude of effect on the reduction of *para* in our data. When *para* functions as a directional marker it strongly favors reduction (.70), as compared to the instances where the preposition serves to indicate goal (.45), temporality (.32), or other (.15) functions.

Next, we observe the results pertaining to the *para* + WORD collocation which reveal a connection between the most frequently co-occurring elements and the reduction of *para*. That is, when the word following *para* was among the most frequent words co-occurring with *para* (e.g. the 382-493 and 34-231 token categories), the weights indicate that reduction is favored, (.57 and .50, respectively). When the following word was among the least frequent words co-occurring with *para* (e.g. the 5-29 and 1-4 token categories), reduction was disfavored, (.48 and .44, respectively). Finally, regarding the WORD + *para* collocation, we see that reduction was favored in the more frequent combinations (.51), while reduction was slightly disfavored in the least frequent combinations (.49).

Table 4. Rates of reduction for *para* in Caracas Spanish by context

Total N		5310		
% reduced		38.5%		
corrected mean		0.363		
	Weight	**% Reduced**	**% of data**	**N**
Preposition Meaning				
Directional	.70	58.1%	32.4%	1720
Goal	.45	32.2%	55.5%	2947
Temporal	.32	22.0%	4.9%	259
Other	.15	9.6%	7.2%	384
Range	*55*			
***para* + WORD Frequency**				
382-493 tokens	.57	46.0%	23.8%	1261
34-231 tokens	.50	39.8%	36.9%	1953
5-29 tokens	.48	35.9%	20.0%	1061
1-4 tokens	.44	29.7%	19.3%	1021
Range	*13*			
WORD + *para* Frequency				
4-109 tokens	.51	42.5%	56.9%	2228
1-3 tokens	.49	35.9%	43.1%	1690
Range	*2*			

log-likelihood = -3259.977; χ^2/ cell =1.6264 ; p $\leq$ 0.025

5. Discussion

In the previous section, we showed that frequency was one of the most important factors conditioning the reduction of *para*. At the same time, reduction was favored when the preposition functions as a directional marker. As the word following *para* increased in frequency of co-occurrence, so did reduction of the preposition to *pa*. This was also the case for reduction as it related to frequency of co-occurrence of the immediately preceding word. These results provide preliminary support for our hypothesis that frequently-occurring multi-word units in discourse should be stored in the lexicon

(Bybee 2001), easing cognitive processes in the form of automated retrieval and reproduction, thus predicting the reduction of *para* in cases where the degree of constituency is highest.

Turning now to the finding which revealed that the *para* + WORD context produced a stronger favoring effect for reduction as compared to the WORD + *para* context, it will be noted that there were fewer opportunities for the formation of collocations in the latter case. That is, the maximum frequency of a *para* + WORD collocation was 493, whereas the maximum frequency of a WORD + *para* collocation was only 109. Moreover, there were twice as many unique collocations for WORD + *para* than for *para* + WORD. This means that the WORD + *para* context is composed of many infrequent collocations, few of which are of sufficient frequency to attain autonomy of storage. The *para* + WORD context is composed of a comparatively small number of more frequent collocations, many of which are of sufficient frequency to attain autonomy of storage. The difference in the strength of the two effects can, therefore, be attributed to these differing numbers of opportunities for the lexicalization of the collocations to occur. Despite these differences, the effect seems to hold for both collocations.

It is also interesting to note that in the *para* + WORD context the frequency effect was evidenced across all categories in a gradient manner. That there was a gradient effect for the *para* + WORD context is, in and of itself, a considerable finding. That is, while the majority of studies considering frequency treat this phenomenon as a binary variable, our results confirm the stance taken by Bybee and Scheibman (1999) in that frequency effects should be seen in a scalar manner in direct correlation with the strength of the constituents. In our data a scalar effect for *para* reduction was evidenced dependent upon frequency of co-occurrence of the words immediately following the preposition, thereby providing evidence for variable scalar degrees of constituency. Thus, elements that are most likely to appear together in discourse are more likely to be stored as a unit in the lexicon. Equally as interesting, and important for our analysis, is the fact that semantic and formal similarity leads to more reduction in *para*. This was seen in the fact that a combination of directional meaning of the preposition with a co-occurring adverb that begins with an /a/ produced a greater rate of reduction above the overall frequency effect.

Lastly, it will be recalled that Bentivoglio et al. (2005) analyzed a total of 1599 tokens, from 48 speakers, excluding what they believed invariable constructions in their data, such as *para + aqui*. In our analysis we included all tokens, from all 160 speakers in the corpus, with the specific goal of explaining those cases that these authors excluded from their envelope of variation, given that our preliminary inspection of the corpus revealed variable production of these specific collocations. Our results add to those of Bentivoglio and colleagues with an analysis that considers frequency effects. Our paper contributes to the literature on usage-based models by providing evidence that reduction of *para* and the degree to which it forms a constituent with co-occurring elements is driven by frequency of co-occurrence.

6. Conclusion

The findings of the present investigation reveal that the frequency of the collocation is a key factor in predicting reduction of the form *para*. The results revealed that frequently used combinations involving *para* are more likely to reduce than less frequent ones in the corpus of 160 native-born residents of the Venezuelan speech community observed. Further, it was shown that frequency of use is a better explanation for the reduction of *para* than more traditional variables such as phonetic context. The frequency of the *para* + WORD collocation had the strongest effect in the data, evidencing gradiency and higher rates of reduction directly in tandem with the frequency of co-occurrence of the following word. When considering WORD + *para* collocations, the effect was weaker. This was due to the fact that the *para* + WORD context evidenced collocations of much greater frequency than the WORD + *para* context. This result provides evidence to support the hypothesis that frequent combinations are accessed as autonomous units, due to their tighter degree of constituency, as in the case of *I don't know* in American English (Bybee and Scheibman 1999). Furthermore, we provided evidence that type frequency best explains the correlation between the high likelihood of reduction of *para* in those contexts where the following adverb has a meaning associated

with directionality. Our paper adds to the growing body of research following Bybee's (2001, 2006, 2010) usage-based model of language, holding that mental representations of linguistic categories are shaped by our linguistic experiences.

References

Alba, Matthew C. 2006. Accounting for variability in the production of Spanish vowel sequences. Selected Proceedings of the 9th Hispanic Linguistics Symposium, ed. by Nuria Sagarra and Almeida Jacqueline Toribio, 273-85. Somerville, MA: Cascadilla Proceedings Project.

Bentivoglio, Paola, Kristel Guirado, and Guillermina Suarez. 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 and Mercedes Sedano. 1993. Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana. *Boletín de Lingüística* 8. 3-35.

Bentivoglio, Paola, and Mercedes Sedano. 2011. Morphosyntactic variation in Spanish speaking Latin America. *The Handbook of Hispanic Sociolinguistics*, ed. by Manuel Díaz-Campos. Oxford: Blackwell.

Boyland, Joyce Tang. 1996. Morphosyntactic change in progress. A psycholinguistic approach. University of California dissertation.

Brown, Earl K. 2008. *A usage-based account of syllable- and word-final /s/ reduction in four dialects of Spanish*. University of New Mexico dissertation.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2002. Phonological evidence for exemplar storage of multiword sequences. *SSLA* 24. 215-221.

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

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan and Joanne Scheibman. 1999. The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37-4, 575-596.

Cedergren, Henrietta. 1973. The interplay of social and linguistic factors in Panama. Cornell University PhD Dissertation.

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

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. 2006. La adquisición de la estructura fonológica y de la variación sociolingüística: un análisis cuantitativo del debilitamiento consonántico en el habla de niños caraqueños. *Haciendo Lingüística*: *Homenaje a Paola Bentivoglio*, 61-75. Caracas: Universidad Central de Venezuela.

Díaz-Campos, Manuel. 2009. *Acquisition of phonological structure and sociolinguistics variables: A quantitative analysis of Spanish consonant weakening in Venezuelan Children´s speech*. Saarbrücken, Germany: VDM Verlag Dr. Müller.

Díaz-Campos, Manuel and Michael Gradoville. 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. Luis A. Ortiz-López, 224-238. Somerville, MA: Cascadilla Proceedings Project.www.lingref.com, document #2490.

Díaz-Campos, Manuel, and Mary Carmen Ruiz-Sanchez. 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.

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.

Garcia, Maryellen. 1979. *Pa(ra)* Usage in United States Spanish. *Hispania* 62.1: 106-114.

Guirado, Kristel. 2007. ¿*Pa donde va a agarra?*: la alternancia para-pa en tres comunidades de habla venezolana. *Interlinguistica* 17. 455-464

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

Hooper, Joan B. 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, Alan Bell, Michelle Gregory, and William D. Raymond. 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.

Pagliuca, William, and Richard Mowrey. 1987. Articulatory evolution. *Papers from the 7ᵗʰ International Conference on Historical Linguistics*, ed. by Anna Giacalone Ramat, Onofrio Carruba, and Giuliano Bernini, 459-72. Amsterdam: John Benjamins.

Stemberger, Joseph Paul., & Barbara Bernhardt. 1999. The emergence of faithfulness. In B. MacWhinney, ed. *The emergence of language*, 417-446. Mahwah, London: Erlbaum.

Willis, Erik W., and Travis Bradley. 2008. Contrast maintenance of taps and trills in Dominican Spanish. Data and analysis. *Selected Proceedings of the 3ʳᵈ Conference on Laboratory Approaches to Spanish Phonology*, ed. by Laura Colantoni and Jeffrey Steele, 87-100. Somerville, MA: Cascadilla Proceedings Project.

Selected Proceedings of the 14th Hispanic Linguistics Symposium

edited by Kimberly Geeslin and Manuel Díaz-Campos

Cascadilla Proceedings Project   Somerville, MA   2012

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:

Díaz-Campos, Manuel, Stephen Fafulas, and Michael Gradoville. 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. Kimberly Geeslin and Manuel Díaz-Campos, 75-87. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2657.