The Value of Frequency as a Linguistic Factor: The Case of Two Dialectal Regions in the Spanish Speaking World

Manuel Díaz-Campos¹ and Carmen Ruiz-Sánchez²
Indiana University¹ and Michigan State University²

1. Introduction

Sociolinguistic studies have traditionally included multiple linguistic factors in their efforts to find patterns of language variation, but they have neglected the role of lexical frequency in such variation for the most part. The results of some recent investigations, however, reveal that the inclusion of lexical frequency as an additional linguistic factor can help achieve a better understanding of language variation and change (Bybee 2002, Brown 2004, Alba 2005, Díaz-Campos 2005, 2006, Ruiz-Sánchez 2006, 2007). Language use is a key factor in identifying patterns that lead to language change and can be useful to understand relationships among linguistic units that are not captured by current models in linguistic theory. Frequency is taken as evidence of patterns that might be subject to automation and changes in mental representations. Bybee (2001: 11) states: “If sound changes are the result of phonetic processes that apply in real time as words are used, then those words that are used more often have more opportunity to be affected by phonetic processes.” The case of variable phenomena is particularly relevant since variation has been overlooked by formal theories. One of the advantages of exemplar theory is that it allows accounting for phonetic detail as part of lexical representations. This type of approach is fundamental to understand variable production in the input and to explain how usage promotes the spreading of changes along with sociolinguistic variables such as age, gender, socio-economic class, just to name the most commonly studied variables in the field.

The goal of this paper is to demonstrate the value of frequency as an independent variable in explaining sociophonological variation across language varieties, and to encourage sociolinguistic researchers to enquire about the role of usage in explaining language change. Specifically, we study Andalusian and Venezuelan Spanish with the purpose of comparing the same phenomenon in two dialects that are historically related, but that given a great geographical distance show certain independence in the development of such phenomenon. The linguistic change in question is syllable-final /-r/ deletion (e.g. [kanta] instead of [kantar] cantar ‘to sing’), which is attested in both varieties and is described as common to both of them as well.

2. Previous research
2.1. Deletion of /-r/ in Andalusian Spanish

Previous research on /-r/ deletion in Andalusian Spanish is very limited. There are some general studies that describe the social and regional distribution of the main characteristics of Andalusian varieties, and they often include some information about deletion of /-r/ (Alvar 1979, 1988, 1996, Cano 2000, Carbonero 1982, 1992, 2003, Gordón & Ruthstaller 1992, Jiménez 1999, Mondéjar 2001, Narbona, Cano & Morillo 2003, Uruburu 1996, Zamora Vicente 1970, among others). These studies agree that deletion of /-r/ is widely spread throughout the Andalusian territory, being more frequent in word-final position, in the sequences /-rl-/ and /-rn-, and among less educated speakers. García Marcos (1992) takes into account some additional linguistic and social factors in his study of the stratification of /-r/ in the Spanish of the coast of Granada, and concludes that this process also tends to occur in pre-lateral and pre-obstruent contexts, in infinitives, in the oldest generation and in female speech.
More recently, Ruiz-Sánchez (2006, 2007) provided a usage-based account of the phenomenon in Andalusia, in which multiple linguistic and social factors were taken into account. A multivariate analysis of the data revealed that /-r/ deletion is favored by all following phonetic contexts other than a following obstruent, by infinitives, word-final position, and stressed syllables. As for its social distribution, the phenomenon was found to be more common among the youngest generation, the lower education groups and female speakers. These results were tentatively interpreted as an indication that /-t/ deletion, an apparently stable variable phenomenon, could be giving way to a change in progress in the community of Alcalá de Guadaíra (Seville). On the other hand, the analysis of word frequency provided evidence that there is a pattern in the diffusion of this variation or phonological change, which is affecting high-frequency words before low-frequency ones. According to Bybee (2002), this type of phonetic change can be characterized as phonetically conditioned and gradual (both phonetically and lexically), and syllable final /t/ deletion in this Andalusian variety shows this pattern of lexical diffusion. The findings about the effect of frequency on /-t/ deletion add to the body of studies that have proved the effect of language use in shaping language structure, and provide important insights into the nature of linguistic variation and change.

2.2. Deletion of /-r/ in Caracas Spanish

General descriptions about Venezuelan Spanish point out that liquids in syllable-final position are subject to alternation and deletion. Lipski (2005) maintains that lateralization and rotacismo (i.e. production of /-l/ instead of /-r/) are rare in Caracas, while deletion is much more common. He points out that lateralization and rotacismo are sociolinguistically marked and that they are found in the lower socio-economic groups. A corpus of 36 recordings is examined by D’Introno, Rojas and Sosa (1979). Their findings indicate that liquids show a variable production including intermediate productions and total deletion. Regarding deletion of syllable-final /-r/, D’Introno et al. (1979) found that omission of this sound is more likely to happen in words such as porque ‘because’, por ‘for’, and infinitives. The sociolinguistics analysis shows that men and lower class speakers favored deletion. Consistent with the observation made by Lipski (2005), D’Introno et al. found that /-r/ produced as /-l/ is rare and that it is only found in lower socio-economic class speakers and especially men. An interesting finding is that the production of /-l/ as /-r/ is also infrequent, but favored by women, not by men. They also presented a generative account in which rules are proposed for the phenomena as if these processes were independent. D’Introno et al. argued for two types of rules: (1) weakening rules, and (2) rules of change. They point out that the weakening rules affecting /-r/ are more frequent in the Caracas corpus. The formal analysis proposed does not fulfill the simplicity and elegance required in this type of framework. Including frequency and usage-based approaches to the analysis of this phenomenon would allow unifying the account of syllable-final liquids as interdependent variations as supposed to independent ones.

The motivation to provide a novel analysis using exemplar theory was carried out by Díaz-Campos (2005). He examines syllable-final /-t/ deletion in order to determine the role of frequency in the diffusion of this phonological change in Venezuelan Spanish using the corpus Estudio Sociolinguístico de Caracas (1987), which comprises speech samples of people born and raised in the capital city of Venezuela. The results, based on 7,200 tokens, show a distribution of 30.2% deletion, 2.5% lateralization and 67.4% retention. Phonetic context, grammatical category, age, socio-economic class, and gender were the group factors statistically predicting syllable /-r/ deletion. All contexts with the exception of a following obstruent favored deletion while older, male and lower socio-economic class speakers deleted more than other groups. The frequency analysis shows that syllable-final /-r/ presents a pattern that affects high-frequency words earlier than low-frequency ones. Deletion is more likely to happen in infinitives and conjunctions: the most frequent categories in our corpus. The evidence presented in the previous literature reviewed warrants further analysis of syllable-final /-r/ using exemplar theory since there are indications of its positive effect in understanding patterns of language change.

This paper presents a comparative analysis of syllable-final /-r/ in Andalusian and Venezuelan Spanish with the purpose of determining the role of frequency in accounting for the development of the same socio-phonetic phenomenon in two different dialects. The following research questions
guided our analysis: (1) What are the linguistic and social factors conditioning syllable-final /-r/ deletion in both Andalusian and Venezuelan Spanish? (2) What is the role of frequency in predicting the spreading of the change in both Andalusian and Venezuelan Spanish?

The next section is concerned with the methodology of the study. We provide information about the corpora, the dependent and independent variables included, and the statistical analysis performed on the data.

3. Methodology and Analysis

The Andalusian corpus consists of 36 sociolinguistic-style interviews conducted with native speakers from the Andalusian town of Alcalá de Guadaíra in Summer 2006, who were evenly grouped in terms of education (elementary, high school, and college university), age (21, 34, 35, 55, and 56 and older), and sex (male and female). For Caracas, 36 speakers were selected from the corpus Estudio Sociolingüístico de Caracas (1987) with equal representation of socio-economic level (upper class, middle class, and lower class), age (14, 29, 30, 45, 61 and older), and sex (male and female). The dependent variable is syllable-final /-r/ production, which has three distinct variants: (1) retention, (2) deletion, and (3) lateralization. Six independent variables were included to examine Andalusian and Venezuelan Spanish:

1. Following phonetic context. The effects of the immediate sound following syllable-final /-r/ were coded for the analysis. Sounds were classified as belonging to one of these categories: obstructant, sonorant, and vowel. In the case of absence of sound we coded a pause.
2. Grammatical category. The lexical items, where instances of syllable-final /-r/ were found, were coded in the following categories: infinitives, conjunction, noun, adverb, adjective, and preposition. Grammatical category is one of the traditional independent variables included in variationist analysis and was also considered by D’Introno et al (1979).
3. Following morphemic unit. This independent variable is concerned with whether syllable-final /r/ is followed by a clitic as in comerlo ‘to eat it’ vs. comer ‘to eat’. Is it the case that instances of syllable-final /r/ in the clitic context are more likely to be deleted? This variable includes three variants (1) infinitive + cltic, (2) infinitive, (3) other cases.
4. Age. Age groups are different for the Andalusian and the Caracas corpora. These corpora were collected at different periods of time and with distinct goals, which explains their unique designs. However, based on the data we have, we propose comparisons among groups that are in the same age range. The Andalusian corpus comprises three age groups: (1) 21-34 years old, (2) 35-55 years old and (3) 56 or more years old. The Caracas corpus includes the following groups: (1) 14-29 years old, (2) 30-45 years old, (3) 61 or more years old.
5. Sex. Participants were divided by their biological sex in male and female.
6. Socio-cultural level: the Andalusian corpus classified speakers according to level of formal education into three groups: (1) elementary, (2) secondary, (3) university. The Caracas corpus has a more complex definition of socioeconomic class by taking into account level of salary, education, salary and level of education of the father and mother, residential area, etc. (see more details in Bentivoglio and Sedano 1993). Three groups were included in the present study: (1) upper class, (2) middle class, and (3) lower class.

The second part of the analysis is concerned with the effect of frequency on /-r/ deletion in Andalusian and Caracas Spanish. Frequency is defined in two ways in this investigation. On the one hand, lexical or word frequency was determined according to the number of occurrences of lexical items containing syllable-final /r/ in the corpora. Words that occurred 80 times or more in the respective corpus were considered high-frequency items while the rest were coded as low-frequency words. Based on previous investigations (Bybee 2001, Bybee 2002, Brown 2004, Alba 2005, Díaz-Campos 2005, 2006, Ruiz-Sánchez 2006), we predict that syllable-final /r/ is a socio-phonetic phenomenon that is classified as physiological motivated change that spreads from more to less frequent lexical items. The second way of including frequency was to consider the number of cases of retention and deletion of syllable-final /-r/ according to grammatical category. This can be taken as a way to include type frequency, which means finding the patterns that occurred more in the data and that facilitate the spreading of the change.
The data were statistically analyzed with Goldvarb 2001 in order to determine the relative effect of the factors influencing /-r/ deletion in either language variety considering each independent variable and its effect in the dependent variable. We combined the social variables of age and gender to show the behavior of males and females according to age. Our prediction is that despite the geographical separation and lack of contact between these two speech communities, they will present similarities with respect to deletion of /-r/, especially in the predominance and consistency of the frequency effect.

4. Results and Discussion

This section is a summary of the effects that the multiple linguistic and social factors analyzed in this investigation have in both Andalusian and Caracas Spanish. We will start with a brief discussion of the linguistic manifestation of /-r/ deletion in these two varieties, we will then compare the way this phenomenon is socially spreading in the two speech communities studied here, and we will finish with an examination of the value of frequency in explaining variable deletion of /r/ in these two dialects.

4.1. Following phonetic context

As it can be seen in Figure 1, deletion of /-r/ tends to occur in similar phonetic contexts in both Andalusian and Caracas Spanish. The phenomenon is favored by all the following contexts except for a following obstruent. The results regarding phonetic conditioning reveal that in both varieties the change follows the same path. Based on the fact that most of the contexts are favorable to increased levels of deletion, we might suggest that deletion has spread to all contexts in both dialects; hence the apparent lack of clear phonological patterns.

![Figure 1: Probability of /-r/ deletion according to the phonetic context.](image)

4.2. Grammatical Category

Both Andalusian and Caracas speakers tend to delete /-r/ in infinitives the most. The results for this linguistic factor confirm previous tendencies found in the literature (Cedergren 1973, Díaz Campos 2001, 2005, D’Introno, Rojas & Sosa 1979, García Marcos 1992, López Morales 1983, Poplack 1986, Rojas 1988, Uruburu 1996). As for the other grammatical categories, Caracas speakers strongly favor the phenomenon in conjunctions as well, and strongly disfavor it in all the other grammatical categories. Andalusian speakers disfavor deletion in all the grammatical categories other than infinitives but, as it can be seen below, the phenomenon seems to be more spread out in this language variety (Figure 2). For instance, in adjectives and prepositions the probabilities of deletion
are both negative for Andalusian and Caracas Spanish, but Andalusian Spanish has higher probabilities in comparison with Caracas Spanish. This result indicates that it is more likely to find deletion in those paradigms in Andalusian Spanish.

![Probability of /-r/ deletion according to grammatical category](image)

Figure 2: Probability of /-r/ deletion according to the grammatical category.

4.3. Following Morpheme

In Spanish infinitives can have clitic pronouns attached to them. When the presence or absence of a following morpheme is taken into account, we see that this factor does not really have an effect in /-r/ deletion in either Andalusian and Caracas Spanish, since this process is favored in infinitives regardless of the presence or absence of a clitic (Figure 3). In the case of Andalusian speakers, we even observe a stronger tendency towards deletion in infinitives with no clitics, which to some extent contradicts previous observations in the literature that has stressed the high frequency of this phenomenon in the sequences /-rl-/ and /-rn/ that result from the combinations of infinitives with a clitic. Other grammatical categories, coded as “other” for this factor group, disfavor deletion in both dialects, as it has already been pointed out.

![Probability of /r/ deletion according to following morphemic unit](image)

Figure 3: Probability of /-r/ deletion according to the following morphemic unit.
4.4. Social distribution

When considering the social factors of age and gender, deletion of /-r/ was found to present the following distribution in Andalusian Spanish:

![Figure 4: Probability of /-r/ deletion according to age and gender in Andalusia.](image)

The results indicate that this phenomenon is favored by all the speakers except for those in the middle-aged men and older women groups. The fact that younger speakers favor deletion reveals the vitality of the phenomenon in Andalusia and points to the possibility that this variable phenomenon, thought to be a case of stable variation, may be becoming more frequent in the speech of the younger generations as a result of a possible language change.

This process manifests itself quite differently in the Caracas community (Figure 5). In Caracas Spanish, deletion is more likely to happen among the oldest generation and young males, while the other groups show a conservative pattern, as illustrated below. In other words, the process seems to have become stabilized.

![Figure 5: Probability of /-r/ deletion according to age and gender in Caracas](image)
4.5. Social value

The third social factor considered in this investigation was education for the Andalusian corpus and social class for the Caracas one, which allows gaining some information about the social value attached to this same process in two separate speech communities.

![Figure 6: Probability of /-r/ deletion according to social class in Caracas.](image)

Figure 6 reveals that lower socioeconomic class speakers in Caracas favor deletion, while middle and upper socioeconomic groups disfavored it. As for Andalusia, the figure below shows that speakers in the lower educational groups are more likely to delete /-r/ than those with university education.

![Figure 7: Probability of /-r/ deletion according to education in Andalusian Spanish.](image)

These findings indicate that /-r/ deletion has a similar social value in Andalusian and Caracas Spanish. Being more common among the lower social class in Caracas (Figure 6) and the lower educational levels in Andalusia (Figure 7), we can argue that this process is associated with the vernacular variety used in colloquial speech in both communities.
4.6. Frequency effects

We now turn our attention to the frequency analysis of the data in both varieties under consideration in the present paper. We have taken into consideration lexical frequency and measured levels of deletion and retention. The results of the word frequency analysis are illustrated below.

Figure 8: Deletion according to word frequency in Andalusian and Caracas.

It has been observed that phonetic change develops more rapidly in lexical items with high frequency. Bybee (2001:11) further argues that “this effect is particularly noticeable in grammaticizing elements or phrases that undergo drastic reduction as they increase in frequency.” Figure 8 shows that syllable-final /r/ deletion actually increases in high-frequency tokens, while it is disfavored by low-frequency ones. These results indicate a pattern of lexical diffusion typical of physiological motivated phenomena. This means that syllable-final /-r/ is a phonetic change that might be the result of automation in highly used items and sequence of items.

Figures 9 and 10 show deletion cases according to the frequency of grammatical category. This is another way of observing the effects of frequency of usage in different paradigms.

Figure 9: Deletion according to frequency of grammatical category in Andalusia.
As can be seen in Figure 9, the most common grammatical category containing syllable-final /-r/ deletion in Andalusian Spanish is the infinitive where 1,251 were found.

![Number of cases of deletion according to grammatical category in Caracas](image)

Figure 10: Deletion according to frequency of grammatical category in Caracas

The number of cases of deletion in Caracas was 2,171. From that total, 1,548 were instances of deletion in infinitives. There were also frequent instances of deletion in the next most frequent category: conjunctions.

We have shown that, even though the same phenomenon in two different dialects may differ in other aspects of its sociolinguistic manifestation, such as its rate of production in different generations and genders, frequency can still account for the spreading of the phenomenon across dialects: it will affect high-frequency words first and to a greater extent, regardless of whether the phenomenon is manifested as a case of change in progress (Andalusia) or a situation of stable variation (Caracas). Recall that younger speakers favored deletion in Andalusia, while younger informants disfavored it in Caracas.

Frequency also accounts for the apparent irregularity among deletion and other realizations in syllable-final position better than formal analysis. While a rule-based analysis (see D’Introno et al. 1979) cannot explain why rates of deletion are higher in infinitives and conjunctions, our analysis can capture how frequency of usage determines the development of the phenomenon. The fact that both Andalusian and Caracas show similar patterns reveals that our account is consistent across varieties. These findings prove the advantages of including word frequency as an additional linguistic factor in variationist research as it helps understand the motivations for language variation and change.

5. Conclusions

Despite the geographical separation and lack of contact between these two language varieties, Andalusian and Caracas Spanish present some important similarities with respect to deletion of /-r/, specially in the way the phenomenon is spreading through the different linguistic contexts and through the lexicon. The results indicate that deletion of /-r/ tends to occur in the same phonetic contexts in both varieties and that the phenomenon has an analogous behavior in the grammatical category of infinitives. Furthermore, lexical frequency was found to display parallel patterns in both dialects as deletion of /-r/ is more likely to affect high-frequency words than low-frequency ones. Finally, the examination of education and social class in Andalusia and Caracas respectively revealed another similarity: deletion of /-r/ has low prestige in both varieties.
The most relevant difference between these two dialects was found in the social manifestation of this phenomenon, more specifically in the rate of deletion across the different age and gender groups. While deletion of /-r/ shows more vitality in the youngest generation and in female speakers, suggesting that this process may still be advancing in Andalusian Spanish, the results for the Caracas corpus show that the situation may have become stabilized in Caracas Spanish.

From these results we conclude that even though the same phenomenon in two different dialects may differ in some aspects of its sociolinguistic manifestation, such as its rate of production in different generations and genders, it can show a consistent pattern in its linguistic conditioning and, more importantly, in the effects of word frequency. In this paper we have demonstrated that frequency can account for the spreading of a phonological phenomenon across dialects: it will affect high-frequency words first and to a greater extent, regardless of whether the phenomenon is manifested as a case of change in progress (Andalusia) or a situation of stable variation (Caracas). In both Andalusian and Caracas Spanish, syllable-final /r/ fits the pattern of physiological sound change where spreading goes from more frequent items to less frequent ones, or in Bybee’s words (2002), a phonological process that is both lexical and phonologically gradual. These findings prove the advantages of including word frequency as an additional linguistic factor in variationist research as it helps achieve a better understanding of the forces behind variation and language change.

References


