

Perceptual Categorization of Dialect Variation in Spanish

Manuel Díaz-Campos and Inmaculada Navarro-Galisteo
Indiana University

1. Introduction

Research studies on perception of dialect variation by naïve listeners (e.g. Clopper and Pisoni, 2004a, 2004b, 2004c; Preston, 1993, 2002) have provided valuable information in the last years about how listeners categorize, group and judge different dialects. The so-called indexical properties of speech allow us to perceive not only the talker's gender but also, where the talker is from, his ethnicity, age and even his socioeconomic status. These studies have provided the field of speech perception with new insights from disciplines like sociolinguistics, whose concentration is linguistic variation and change; showing the importance of the long ignored phonetic variability and bringing a new approach to speech perception. In this sense, the members of a speech community talk, judge and interpret in a similar way linguistic variables which let them distinguish its members. Talkers are judged as members or intruders of a dialect community depending on the speech characteristics they do or do not share. Thus, the study of the perceptual categorization of dialect variation permits us to learn more about how members of a speech community perceive, classify and distinguish their own dialect from different ones as well as which dialects they are able to distinguish as different or similar to their own.

Little is known about categorical perception of dialect variation as well as the role of linguistic experience in the perception of Spanish dialects. Most of the previous studies about Spanish dialectology (Alvar, 1996; Canfield, 1981; Lipski, 1994) focus on description of linguistic features that are salient in different areas of the Spanish-speaking world. To our knowledge, there are few investigations analyzing dialectal perception in Spanish. Moreno Fernandez and Moreno Fernandez (1999) conducted the first study, which focuses on the perception of dialectal differences among speakers from Madrid. Alfaraz (2002) analyzed Cuban speaker's attitudes toward other varieties of Spanish in the Florida area. Her analysis also provides information about the role of time of arrival (pre-revolution and post-revolution), race, and socioeconomic class in how Cubans evaluate their own varieties of Spanish. With this in mind, the present paper explores perceptual categorization of dialect variation in two groups of Spanish speakers from Latin America and Spain. Most recently, Boomershine (2006) investigated the perception of several phonological variables by Mexican and Puerto Rican listeners.

With this pilot study we attempt to expand our knowledge with respect to how listeners perceptually categorize different Hispanic dialects, as well as to explore the role that linguistic experience and familiarity with a certain dialect play in this process.

2. Previous Literature

First, we begin with a general explanation of indexical features in the speech signal. This is an important issue since regional variation is considered an indexical property of speech. We also present, briefly, findings in the area of speech perception according to which specific characteristics of the talker's voice enable better perception. Secondly, we review research about dialect perception in English and other languages. Finally, a description of investigations with Spanish dialects is presented.

Abercrombie (1967) defined indexical properties of the speech signal as those which provide an index to some attribute of the talker, such as membership in a group (gender, regional dialect), talker specific idiosyncrasies, which allow us to identify familiar talkers, and context specific attributes of the talkers, such as current emotional state. Thus, the speech signal not only carries physiological sources

of variation (male-female) but also sociolinguistic sources of variation (socioeconomic status, region of origin, ethnicity, etc). In our particular case, we focus on the study of group membership as reflected in dialect variation. In the last two decades, extensive research on talker variability in speech perception and word recognition (Goldinger et al., 1991; Mullenix et al., 1989; Nygaard et al., 1994; Pisoni and Lively, 1995) has shown that listeners encode specific information about a talker's voice. Listeners are sensitive to talker variability and they are able to encode indexical features as well as to store them in the long-term memory (Pisoni, 1997). The implication of this type of findings is that idealized models where linguistic features were separated from properties associated with speaker's voice did not reflect accurately what information the speech signal carried and what the listeners perceived.

During the last decades, it has been documented that some properties are easier to categorize than others for naïve listeners. Lass et al. (1976) reported that listeners were 96% accurate in categorizing unfamiliar speakers according to gender, after listening to isolated vowels. Categorization based on social dimensions seems to be more difficult. Preston (1993, 2002) studied dialect categorization on a north and south continuum by naïve adult listeners from Michigan and Indiana. Preston (1993) found that although listeners were not able to know where the talkers were from, they were able to distinguish if talkers were from the north or the south. Preston (2002) provides a possible explanation to the results of his previous study, according to which the listeners from Michigan made their identifications following "correctness", while people from Indiana followed the criterion of "pleasantness".

Van Bezooijen and Gooskens (1999) investigated dialect identification of Dutch and British varieties. They focused on the importance of segmental vs. prosodic features on dialect categorization and found that segmental features were more important for dialect recognition. In their experiments, Dutch listeners performed with around 35% accuracy and British listeners were able to identify talkers from different British regions about 52% of the time.

Williams et al. (1999) asked male adolescent from Welsh to identify other male talkers by dialect in an eight alternative forced-choice task. The results showed 30% of accuracy. These findings also indicate that subjects from the same region would be more-likely to identify accurately the target dialectal region (45% accuracy).

In recent studies about dialect variation, Clopper and Pisoni (2004) investigated how listeners identified dialectal regions, and what acoustic-phonetic cues were useful to them during dialect categorization. They used sentence-length utterances from 11 male speakers in their twenties from the TIMIT Acoustic-Phonetic Speech Corpus, all of them from different regions of the United States. Listeners were asked to listen to two different calibration sentences and eight novel sentences, and categorize each talker into one of six geographic regions, using a touch screen display of continental United States. The results showed a 30% accuracy on dialect categorization. On the other hand, the results of the categorization experiment revealed the capability of listeners to classify speakers using a three-cluster structure formed by the following regions: New England, West, which includes West, North, and North Midland; and South, which includes South, and South Midland. Across experiments, the authors also found that listeners attend to certain phonetic cues in the speech signal that help them affiliate the talker to a certain dialect, even when explicit instruction is not available.

These previous studies show that naïve listeners encode indexical information about unfamiliar talkers. They group them into dialect regions and they follow certain acoustic-phonetic cues that guide them during their decision. However, they also show that although naïve listeners are able to perceptually categorize different dialects, their accuracy performance is still low compared to the categorization of other indexical properties of speech like gender identification.

2.1. The role of residential history and linguistic experience

Effects of residential history and linguistic experience have been documented in various studies. Clopper and Pisoni (2004a) investigated the effects of residential history in dialect categorization, using two groups of listeners; one consisting of listeners that had lived only in Indiana, the "homebodies", and another group consisting of listeners who had moved around the country, the "armybrats". They found that the "armybrats" performed better in the dialect categorization task than the "homebodies", confirming the relevance of linguistic experience based on previous developmental history in the perception of dialect categorization.

The most recent study about the role of amount of experience and region of origin by Baker et al. (in press) also corroborates the importance of familiarity with a variety of dialects in dialect identification. The results of the study about Utahan or non-Utahan dialect identification showed that those listeners who had been living in Utah for a year or less were not able to differentiate between Utahan and non-Utahan speakers. Like previous studies (Clopper and Pisoni, 2004b; Williams et al., 1999) they also found that region of origin is an important factor in dialect identification, since those listeners from the Utah area and the West were more successful in dialect identification than those from more distant regions (non-Western states).

2.2. *Studies on dialect perception in Spanish*

While there is a growing interest in the investigation of perception of dialect variation in the United States, such research related to Spanish has not been very productive. The great majority of the research has been language attitudinal studies in which researchers provided their participants with questionnaires and then asked them questions related to their attitudes towards the variety they were hearing from a tape. In the research article, *Madrid perceptions of regional varieties in Spain*, Moreno Fernandez and Moreno Fernandez (2002) used speech samples and questionnaires given to listeners from the Madrid area. Subjects were provided with a map of Spain, which was divided into the different autonomous regions. Then participants were given a questionnaire in which they were asked to relate on a scale from 1 to 4 if they spoke like the individuals of the regions represented on the map with 1 being 'they speak exactly as I do' and 4 'I find it difficult to understand them'. The results showed that Madrid listeners recognized as 'different from themselves' speakers from bilingual areas of Spain. They also recognized as different from them speakers from the south of Spain and the Canary islands, where the dialects share many features with varieties of Spanish spoken in Central and South America. However, listeners seemed to have more difficulties distinguishing the northern and central dialects as 'distant' from their own. It is worth noticing that responses in the social distance task varied according to gender and age. Thus, social factors might influence the answers of the listeners.

Most recently, Alfaraz (2002) investigated Miami-Cubans attitudes towards several varieties of Spanish; Latin American, Peninsular Spanish and two Cuban varieties (in order to study the perceptions of Cuban Spanish at two different points in time: before and after the Cuban Revolution). The participants were 148 Cubans (male-female) between the age of 20 and 60 and were divided into those who arrived early and those who arrived in later waves of immigration. Following Preston's methodology to elicit evaluative data, participants were asked to rate the pleasantness and correctness of the three dialects. The results showed the generic Peninsular variety to be perceived as the most correct one followed by the Cuban-Pre. The variety spoken in Miami after the Revolution is not considered as pleasant and correct as the variety spoken before the Revolution. Although race and economy seemed to play a role in the results, Alfaraz (2002) pointed out that political ideology was the decisive factor.

Boomershine (2006) studied the perception of three phonological variables by Mexican and Puerto Rico listeners in three different tasks (timed naming, lexical decision and identification task). According to the results, syllable-final /s/ seems to be a relevant acoustic cue in distinguishing Mexican from Puerto Rican Spanish in all three tasks.

3. Current Study

Considering the results obtained in previous studies and taking into account the limited research on dialect perception in Spanish, the goal of this study is to investigate perceptual categorization of dialect variation by listeners from Venezuela and Spain with the purpose of examining the relationship between linguistic experience and the indexical properties of speech associated with dialect variation. For the present article, we examine the following research questions: 1) Are naïve listeners able to perceptually categorize talkers from different Spanish speaking countries? 2) What are the dialectal regions that listeners are able to recognize? How accurate are they? 3) What is the role of linguistic experience in dialect categorization? 4) Does length of exposure to the experimental stimuli influence the accurate categorization of the dialects?

3.1. *Talkers*

Twelve talkers, 6 males and 6 females, from six different Hispanic countries (Spain, Chile, Argentina, Colombia, Costa Rica and Mexico) were selected from the oral corpus Catalogue of Spanish Sounds, provided by Professor Terrell Morgan and recorded in 1999. This corpus was originally conceived for the purpose of teaching Spanish pronunciation and it includes speakers representing different dialectal areas of the Spanish-speaking world. All the speakers are from the capital city of their corresponding country. At the time of the recording all speakers were enrolled as graduate students in different programs at The Ohio State University in Columbus, OH. The age range of all the participants was 25 to 35.

3.2. *Stimulus materials*

The first experiment consisted of a passage which was divided into sentences. A total of 93 sentences were played to our participants. This was done in order to investigate the effect of time of exposure to the experimental stimuli. For the second experiment, the stimulus consisted of the complete passage read by all twelve talkers (see appendix 1). A passage instead of a sample of a sociolinguistic interview was selected in order to control for lexical variation. We wanted listeners to categorize the different dialects based on acoustic-phonetic cues and not on the differences in the lexicon across the different regional varieties of the Spanish language. The passage made reference to the diversity of Spanish dialects across geographic areas. The recordings contained acoustic-phonetic features that were produced with the pronunciation of the dialectal region of the speakers as described in the dialectology literature (see Lispki, 1994). These materials, as explained above, were recorded for a very different purpose and therefore, Venezuelan listeners were not able to be exposed to stimuli from their own variety.

3.3. *Participants*

A total of fifty listeners, males and females, from Spain and Venezuela (20 from Spain and 30 from Venezuela) participated in this experiment. These countries were chosen as starting point in a larger project that in the future pretends to analyze dialect perception in the Spanish-speaking world. Both researchers had contact with the particular dialects selected since one of them is a native of Spain and the other is a native of Venezuela. In the case of the data collected in Spain, 8 listeners took the sentence experiment and 12 listeners took the passage experiment. In the case of the data collected in Venezuela, 14 listeners took the sentence experiment and 16 listeners took the passage task. All of them were monolingual Spanish speakers without a hearing disorder at the time of the experiment. The listeners were between 14 and 65 years of age.

A background questionnaire was given to the listeners in order to collect information on residential history (see appendix 2). In the case of the Spanish listeners, 12 had never been in other Hispanic countries, while, 8 had only been to Mexico. On the other hand, 18 of the Venezuelan listeners had never been abroad and 12 had visited at least one other country of the six included in the present study with the exception of Costa Rica. Thus, in order to analyze the effect of residency on categorization performance, both groups of listeners were divided into two groups based on their experience abroad. Those who had been abroad a month or longer fell in the category of 'experience abroad' while those who had been abroad less than a month or never had been abroad fell in the category of 'no experience abroad'. In addition, unlike other studies done before, we took into account the possibility of contact with speakers from other varieties of Spanish within the listener's country. This is also an important fact of linguistic experience which should not be ignored. Therefore, we further divided our listeners into those with contact and those without contact with other Spanish speakers and the Spanish variety that they speak according to our questionnaires. Most of our listeners had had contact with other Spanish speakers.

3.4. *Procedure*

Listeners were provided with a laptop which showed six different response icons, each labeled according to the name of the particular country. The stimuli were presented over headphones. The

computer screen displayed the six different choices; each country had a number, which the listeners used in order to provide their answer on the computer keyboard.

The listeners heard each of the talkers, presented in random order, once. After listening to the stimulus, the listeners had to select the region, represented by a number, and to provide the answer. No feedback was given during the experiment. Time reaction was recorded after participants pressed the bottom selecting their answer.

4. Analysis

The results of both experiments were analyzed based on accuracy on the categorization of the different dialects, where reaction time was also measured, and based on linguistic experience (further divided into the two categories ‘experience abroad’ and ‘contact’.) As it was mentioned before, unlike other experiments, one of our goals was to learn more about the role of linguistic experience. In order to do that, it was important to take into account those who had been in other countries and contrast their performance with those listeners who had never been abroad. However, linguistic experience should not only be defined as time spent in other countries, like other studies have done before (Clopper and Pisoni, 2004b), but exposure and amount of exposure to a certain Spanish variety. Thus, it is possible that a listener who has never been to Mexico is able to recognize the variety accurately because he has family or friends from Mexico. In this case, this listener should be able to classify correctly this Spanish variety, even though he has never been to that country. For this reason, the results were also analyzed based on the ‘contact’ or ‘no contact’ with Hispanics from the countries presented in the experiments.

5. Results

5.1. Results according to accuracy of dialect categorization in sentence experiment

In this section, we first present the results based on accuracy of identification of the different varieties of Spanish according to each experiment. We take into account reaction time and compare reaction time with accuracy in order to analyze if the results mirror each other. The less time spent on the categorization, the higher the accuracy. Central to our discussion of linguistic experience are also the results related to experience abroad and contact, defined according to the background questionnaire as follows: experience abroad means that participants have traveled to one or several of the areas included in the study. As mentioned before, contact was defined according to the participants’ experience with speakers from other regions whether they are friends or family. We begin with the results of the sentence experiment. Table 1 shows the results of accuracy of dialect categorization for participants from Spain.

Dialectal Region		Accurate	Not accurate	Total
Argentina	N	29	99	128
	%	23	77	
Chile	N	29	91	120
	%	24	76	
Colombia	N	27	101	128
	%	21	79	
Costa Rica	N	32	88	120
	%	27	73	
Spain	N	102	26	128
	%	80	20	
Mexico	N	26	94	120
	%	22	78	
Total	N	245	499	744
	%	33	67	

Table 1: Accuracy of dialect categorization. Sentence experiment. Spain.

First, we explain the results and how they were tabulated. Stimuli for some dialects included 16 sentences (i.e. Argentina, Colombia, Spain), and some 15 sentences (i.e. Chile, Costa Rica, and Mexico). The number of participants taking this task was 8. Taking into consideration that 16 sentences were included for Argentina, Colombia, and Spain and that 8 speakers took the task, give us 128 total stimuli for these areas, whereas the total stimuli is 120 for countries in which 15 sentences were included.

Participants from Spain recognized with success their dialect because we can see that they obtained 80% of accuracy in this particular case. For other regions, levels of accuracy range from 21% to 27%. Although low, all these results are above chance level because random selection would be 17% or below¹. These levels of accuracy are not surprising since the previous studies (e.g. Clopper and Pisoni, 2004a, 2004b, 2004c; Preston 1993, 2002) report similar results that indicate the difficulty participants have recognizing regional indexical properties of speech. Table 2 presents accuracy of dialect categorization in the sentence experiment for Venezuelan participants.

Dialectal Region		Accurate	Not accurate	Total
Argentina	N	78	146	224
	%	35	65	
Chile	N	86	124	210
	%	41	59	
Colombia	N	81	143	224
	%	36	64	
Costa Rica	N	44	166	210
	%	21	79	
Spain	N	123	101	224
	%	55	45	
Mexico	N	48	162	210
	%	23	77	
Total	N	460	842	1302
	%	35	65	

Table 2: Accuracy of dialect categorization. Sentence experiment. Venezuela.

As explained above, stimuli for some dialects included 16 sentences (i.e. Argentina, Colombia, Spain), and some 15 sentences (i.e. Chile, Costa Rica, and Mexico). The number of Venezuelan participants taking this task was 14. Taking into consideration that 16 sentences were included for Argentina, Colombia, and Spain and that 14 speakers took the task give us 224 total stimuli for these areas, whereas the total stimuli is 210 for countries in which 15 sentences were included.

Participants from Venezuela recognized with success stimuli from Spain with 55% of accuracy. They also perceived with more accuracy Chilean, Colombian, and Argentinean varieties. For the other regions, levels of accuracy ranged from 21% to 23%. Specifically, the lowest accuracy is with Costa Rica (21%) and Mexico (23%). It is worth noticing that these results are above chance because random selection would be 17% or below. Venezuelan participants were generally better than Spaniards in recognizing dialectal areas, but still levels of accuracy do not reach more than 55 percent and in average 35%. A chi-square test shows that the difference between both groups of speakers is significant ($P = .00739$). We also have to remember that no stimuli from Venezuela were included in the experiment. If we take out stimuli from Spain, where both Spaniards and Venezuelan, performed

¹ This criterion is based on previous research by Clopper and Pisoni (2006: 204) according to which “chance performance in a six-alternative task is 17%. Chance performance is $1/N$ where $N=6$ in a six alternative task. This criterion was selected because it is directly tied to the design of the experiment, which follows previous studies in the area.”

the best, we can see that Venezuelans were much better in recognizing areas such as Argentina, Chile and Colombia.

In summary, both Spaniards and Venezuelan participants were successful at distinguishing speakers from Spain. In the case of Venezuelans, positive trends are also found for identifying speakers from Chile, Colombia and Argentina. All tendencies found are above chance. Overall accuracy results indicate 35% in the case of Venezuelans and 33% in the case of Spaniards. We now turn to reaction time results. The rationale behind measuring reaction time was to determine the degree of familiarity of listeners with dialect regions and to compare these results with the levels of accuracy in the categorization experiments.

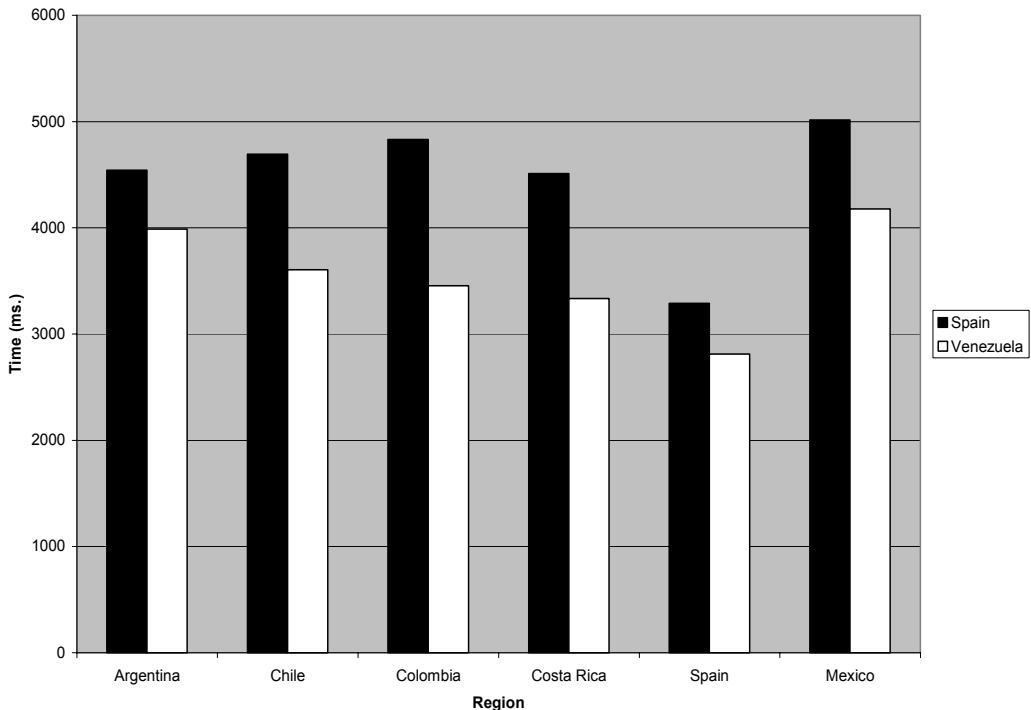


Figure 1: Reaction time in the sentence task (Spain-Venezuela)

As Figure 1 indicates, reaction time measurements were taken from participants completing the task of dialect categorization. Time was measured in milliseconds. These results mirror those of the categorization task. Spaniards spent less time categorizing their own dialect and they needed much more time to categorize the other ones, especially to categorize the Mexican variety.

In comparison with Spaniards, Venezuelan participants spent less time completing the sentence task. We can also see a trend, according to which less time was required for dialects with higher recognition accuracy. For instance, Venezuelans spent less time with stimuli from Spain where levels of accuracy reached 55%. Other cases are less clear, but levels of accuracy were in the low range.

5.1.1. Results of dialect categorization according to contact in sentence experiment

Considering the number of listeners who reported having friends or family members from other dialect regions, accuracy levels were measured. For instance, in the case of Argentina, five Spaniards reported having family or friends from this region, while three other participants reported not having contact. The five listeners were exposed to eighty sentences, while the other three were exposed to forty-eight sentences. In particular, five Spaniards indicated having friends from Argentina, two indicated having friends from Mexico, and one indicated having a friend from Colombia.

Dialectal Region		Accurate Contact	Accurate No contact	Not accurate Contact	Not accurate No contact	Total
Argentina	N	20	10	60	38	128
	%	25	21	75	79	
Mexico	N	8	17	22	73	120
	%	27	19	73	81	
Colombia	N	4	22	12	90	128
	%	25	20	75	80	
Total	N	32	49	94	201	376
	%	25	20	75	80	

Table 3: Accuracy of dialect categorization according to contact. Sentence task. Spain.

Levels of accuracy in dialect recognition improved slightly with contact in all cases with the strongest improvement found in those participants reporting having contact with Mexican Spanish. The overall improvement in accuracy with contact only reaches 5%. This tendency, however, was not statistically significant ($P = .9102$). Linguistic experience with other areas of the Spanish-speaking world seems limited for this particular group of listeners.

Dialectal Region		Accurate Contact	Accurate No contact	Not accurate Contact	Not accurate No contact	Total
Argentina	N	25	48	55	96	224
	%	31	33	69	67	
Chile	N	39	46	36	89	210
	%	52	34	48	66	
Colombia	N	56	33	72	63	224
	%	44	34	56	66	
Spain	N	54	72	26	72	224
	%	68	50	32	50	
Mexico	N	16	27	59	108	210
	%	21	20	79	80	
Total	N	190	226	248	428	1092
	%	43	35	57	65	

Table 4: Accuracy of dialect categorization according to contact. Sentence task. Venezuela.

Venezuelan participants reported having contact with Argentina, Chile, Colombia, Spain, and Mexico. In this case we can see a more broad experience with speakers of other dialects. Five Venezuelans indicated having friends from Argentina, five have friends from Chile, Spain and Mexico, and eight reported to have friends from Colombia. These answers suggest that Venezuelans have more exposure to varieties of Spanish tested in this experiment because a good number of them indicated knowing somebody from all areas except Costa Rica. In other words, linguistic experience with other areas of the Spanish-speaking world includes five out of the six dialects considered in this investigation. Because of Venezuela's geographical position, it is not surprising that more Venezuelans indicated that they have contact with Colombians given the close geographical distance between Venezuela and Colombia as well as the large number of immigrants from Colombia who live in Venezuela. Overall, Venezuelan listeners performed better than Spaniards. This finding was found to be significant by the chi-square test ($P = .00495$). The overall improvement in accuracy with contact reaches 8%. People reporting no contact have an overall accuracy of 35%. We can conclude by suggesting that speakers reporting contact perform better in the identification of dialects.

5.1.2. Results of dialect categorization according to experience abroad in sentence experiment

Dialectal Region		Accurate Experience	Accurate No experience	Not accurate Experience	Not accurate No experience	Total
Mexico	N	8	17	22	73	120
	%	27	19	73	81	

Table 5: Accuracy of dialect categorization according to experience. Sentence. Spain

Only two out of the eight participants from Spain reported having experience abroad. These two informants traveled to Mexico. The results indicate a positive trend according to which stimuli from Mexico was recognized accurately 27% by Spaniards reporting contact in contrast with 19% of accuracy for those who did not travel or had lived in Mexico. A chi-square test reveals that this difference is significant ($P = .00124$).

Dialectal Region		Accurate Experience	Accurate No experience	Not accurate Experience	Not accurate No experience	Total
Argentina	N	5	67	11	141	224
	%	31	32	69	68	
Chile	N	13	72	3	122	210
	%	80	37	20	63	
Colombia	N	11	80	19	114	224
	%	37	41	63	59	
Costa Rica	N	3	31	12	164	210
	%	20	16	80	84	
Total	N	32	250	45	541	868
	%	42	32	58	68	

Table 6: Accuracy of dialect categorization according to experience. Sentence. Venezuela

According to the results, six out of the 14 participants from Venezuela reported having experience abroad. One participant indicated having experience with Argentina, one with Chile, another one with Costa Rica, and two with Colombia. The results indicate a positive trend for Chile and Costa Rica, but such trends are not significant ($P = .22497$). Another point to take into consideration is the fact that with just one or two participants reporting experience we do not have the kind of data needed to make strong claims, but suggest trends in the results found. Still it is valuable to observe that these participants report having more experiences abroad in countries included in this study than Spaniards. Overall results indicate an accuracy of 42% for those reporting experience, those participants with no experience were able to identify talkers from these regions 32% of the time. This overall difference reaches statistical significance ($P = .00071$). We will return to this issue concerning the role of experience in dialect categorization later in the paper, in order to provide some suggestions for future research.

We now turn our attention to the results of the second experiment that consisted of 12 passages from the same dialectal areas included in the sentence task, namely Argentina, Chile, Colombia, Costa Rica, Spain, and Mexico. In this second experiment, speakers had a longer exposure to stimuli, which could potentially improve rates of dialect categorization in the results. Twelve participants from Spain completed experiment 2, whereas sixteen participants from Venezuela completed it. Table 7 presents the results concerning accuracy of dialect categorization in the passage task for participants from Spain.

5.2. Results according to accuracy of dialect categorization in passage experiment

Dialectal Region		Accurate	Not accurate	Total
Argentina	N	11	13	24
	%	46	54	
Chile	N	5	19	24
	%	21	79	
Colombia	N	1	23	24
	%	4	96	
Costa Rica	N	6	18	24
	%	25	75	
Spain	N	21	3	24
	%	88	12	
Mexico	N	6	18	24
	%	25	75	
Total	N	50	94	144
	%	35	65	

Table 7: Accuracy of dialect categorization. Passage task. Spain.

The findings in Table 7 are based on data collected from twelve participants from Spain who listened to two samples from each country. In general, participants from Spain were more accurate with stimuli from Spain and Argentina. With the exception of Colombia, all dialects were identified above chance. Overall, the level of accuracy was 35% for the passage task. For Spaniards, in comparison with the sentence task, we see just 2% of improvement, which is not a significant difference according to the Chi-square test ($P=.88134$). Consistently with the sentence task, once again listeners from Spain were very successful in recognizing their dialect more than any other variety.

Dialectal Region		Accurate	Not accurate	Total
Argentina	N	10	22	32
	%	31	69	
Chile	N	10	22	32
	%	31	69	
Colombia	N	18	14	32
	%	56	44	
Costa Rica	N	11	21	32
	%	34	66	
Spain	N	27	5	32
	%	84	16	
Mexico	N	9	23	32
	%	28	72	
Total	N	85	107	192
	%	44	56	

Table 8: Accuracy of dialect categorization. Passage task. Venezuela.

The findings in Table 8 are based on the answers of 16 participants who listened to two passages per country. As in the case of the sentence task, Venezuelans performed much better than Spaniards in dialect categorization. The two varieties that are particularly well categorized are Spain and Colombia. Costa Rica, Argentina, Chile, and Mexico reached levels of accuracy that are a little better than the levels of accuracy found in the sentence task. All tendencies found here are above chance. Linguistic experience defined as contact and experience abroad in this investigation might be related to improved levels of dialect categorization because these groups of Venezuelan listeners reported to have contact

and experience with the regions tested in this experiment than participants from Spain. Overall level of accuracy in the passage task improved to 44, since Venezuelans performed with 35% accuracy on the sentence task. Even though, this trend of improvement is not significant ($P = .24720$), future research based on a larger sample might corroborate the relationship between exposure and accuracy of dialect categorization.

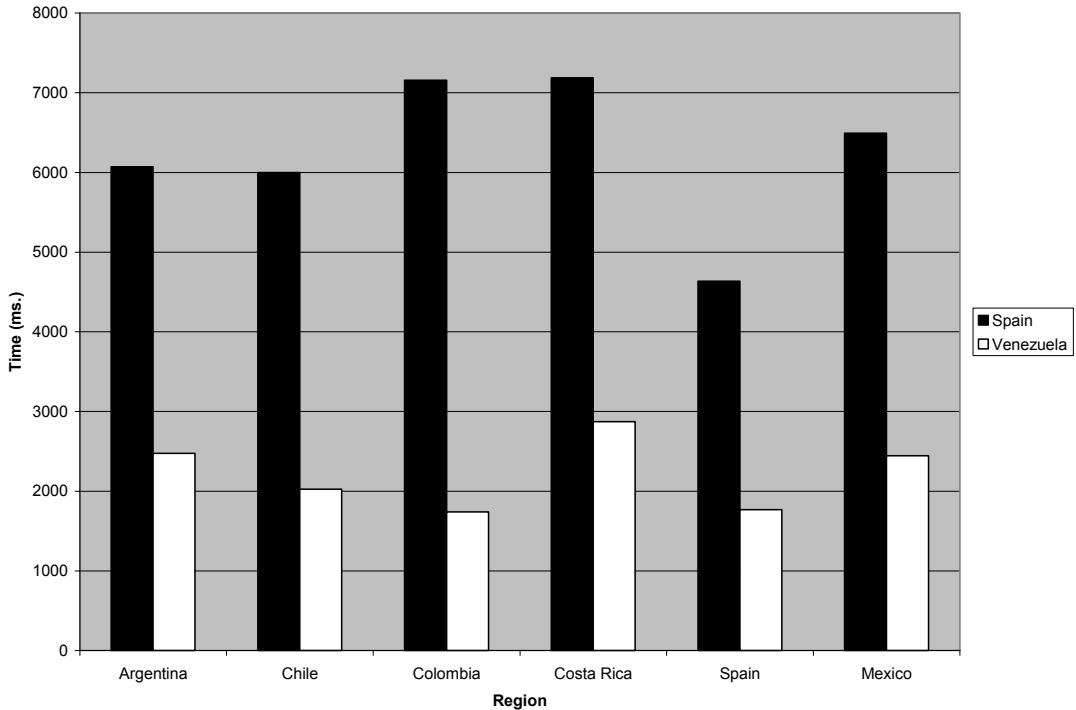


Figure 2: Reaction time for passage task (Spain-Venezuela)

Figure 2 shows reaction time for Spaniards and Venezuelans on the passage task. Recall that for the sentence task Spaniards spent less time in recognizing their own dialect, suggesting that familiarity with this variety facilitates perception. In the particular case of the passage task, once again, participants spent less time with stimuli from Spain. It can be argued that Colombia, and Costa Rica posited some more difficulty because participants spent much more time in categorizing these dialects. In the particular case of Colombia, not only did listeners spend more time, but they were also less accurate. In both cases, one can suggest that lack of experience might have played a role in these results.

In general, Venezuelans spent less time on the perceptual categorization task than Spaniards, suggesting a pattern that is consistent with the idea that this group of Venezuelans has had more linguistic experience with the dialects of Spanish included in this study. We want to point out that in the particular case of Spain and Colombia, Venezuelans spent less time. These are the dialects in which these participants reached the highest levels of accuracy.

5.2.1. Results of dialect categorization in passage experiment according to contact

Dialectal Region		Accurate Contact	Accurate No contact	Not accurate Contact	Not accurate No contact	Total
Argentina	N	3	7	3	11	24
	%	50	39	50	61	
Mexico	N	2	3	2	17	24
	%	50	15	50	85	
Colombia	N	0	2	8	14	24
	%	0	10	100	90	
Total	N	5	12	13	42	72
	%	28	22	72	78	

Table 9: Accuracy of dialect categorization according to contact. Passage task. Spain.

The findings in Table 9 are based on contact defined as having friends or family from any of the six areas included in this investigation. Three participants reported having contact with Argentina, two with Mexico, and four with Colombia. Recall that a total of twelve people took part in this second experiment. Levels of accuracy reached 28% for participants who reported contact, whereas it reaches 22% for informants without contact. A chi-square test indicated that this difference is not significant ($P = .41422$). Nevertheless, the chi-square test revealed that differences found between having or not having contact in the case of Argentina and Mexico were significant ($P = .01265$). This particular result reflects the trend found for two of the dialectal areas with which Spaniards reported contact, suggesting that linguistic experience is a key factor in recognizing indexical features related to geographical affiliation². Future research will have to increase number of participants to observe whether positive trends hold.

Dialectal Region		Accurate Contact	Accurate No Contact	No accurate Contact	No accurate No Contact	Total
Argentina	N	2	8	4	18	32
	%	33	31	67	69	
Colombia	N	9	9	5	9	32
	%	64	50	36	50	
Spain	N	12	15	2	3	32
	%	86	83	14	17	
Mexico	N	2	7	2	21	32
	%	50	25	50	75	
Total	N	25	39	13	51	128
	%	66	43	34	57	

Table 10: Accuracy of dialect categorization according to contact. Passage task. Venezuela

The findings in Table 10 show the results based on contact defined as having friends or family from any of the six areas included in this investigation. Three participants reported having contact with Argentina, seven with Colombia, seven with Spain, and two with Mexico. Recall that a total of sixteen people from Venezuela took part in this second experiment. Levels of accuracy reached 66% for participants who reported contact, whereas it reached 43% for informants without contact. These tendencies were found significant by running a chi-square test ($P = .00178$). This general result reflects

² One might think about the effect of television or of mass media exposure to other varieties of Spanish. This is a controversial topic for the following reasons: (1) investigations in the area of sociolinguistics have not found evidence of any causal influence (Milroy and Milroy, 1985; Trudgill, 1986); (2) even though exposure to media might raise awareness, its concrete effects are secondary since personal contact is primary. Future research might look for ways of designing reliable experimental tools to measure mass media effects in language production and perception.

the trend found for all four dialectal areas with which Venezuelans reported contact. Once again, these findings might suggest that linguistic experience is an important factor in recognizing indexical features related to geographical affiliation.

5.2.2. Results of dialect categorization according to experience abroad in passage experiment

Dialectal Region		Accurate Experience	Accurate No experience	Not accurate Experience	Not accurate No experience	Total
Argentina	N	2	9	0	13	24
	%	100	41	0	59	
Mexico	N	2	4	4	14	24
	%	33	22	67	78	
Total	N	4	13	4	27	48
	%	50	33	50	67	

Table 11: Accuracy of dialect categorization according to experience. Passage. Spain

Regarding experience abroad in the passage experiment, three out of the 12 participants from Spain reported having experience abroad. One participant indicated having experience with Argentina, and two with Mexico. We found few listeners with experience abroad in this particular case. Caution must be taken while analyzing these data. Nonetheless, we observe a positive trend for both Argentina and Mexico, according to which participants reporting experience abroad in a specific dialectal area performed better than those without experience in dialect categorization. Overall results indicate accuracy of 50% for those reporting experience, while accuracy is 33% for those participants with no experience. This difference was found significant ($P = .02167$). Table 12 shows the results of dialect categorization in the passage task for Venezuelans according to experience.

Dialectal Region		Accurate Experience	Accurate No experience	Not accurate Experience	Not accurate No experience	Total
Colombia	N	5	13	3	11	32
	%	63	54	37	46	
Spain	N	3	24	1	4	32
	%	75	86	25	14	
Total	N	8	37	4	15	64
	%	67	71	33	29	

Table 12: Accuracy of dialect categorization according to experience. Passage. Venezuela

Six out of the 16 participants from Venezuela reported having experience abroad. Four participants indicated having experience with Colombia, and two with Spain. Once again, experience abroad is not a common factor in this group of listeners. The tendencies found suggest a positive trend for speakers who have experience with Colombia, while a negative one for those who have experience with Spain. Overall results indicate accuracy of 67% for those reporting experience, while accuracy is 71% for those participants with no experience. These overall tendencies were found not to be significant ($P = .64647$). While we are unable to present definite conclusions, we can suggest that limited experience abroad might be a factor for explaining these unexpected results because this group of speakers reported not traveling abroad or, in few of the cases, traveling for less than a month. In order to provide more evidence we need to expand the sample. We come back to this issue in the discussion section. In summary, participants from Spain with experience with a certain region identify better those areas, while Venezuelans do not show a clear pattern. We turn our attention to the discussion of the findings in the next section.

6. Discussion

These results show the same pattern of performance of previous studies (Clopper and Pisoni, 2006; Clopper, Conrey, and Pisoni 2005; and Clopper and Pisoni 2004a, 2004b, 2003-2004, among others). That is, indexical properties of speech related to regional variation are difficult to identify, since participants in these earlier studies reached 30% of accuracy. Overall, Spaniards and Venezuelans were not very successful in the categorization of other dialects (34% average), confirming findings from previous research. However, it is worth mentioning that listeners are sensitive to dialect perception since they were successful in distinguishing speakers from several regions, reflecting that participants encode indexical properties related to dialect variation. In addition, most of them performed above chance.

Venezuelan listeners performed better than Spanish listeners, categorizing successfully a greater variety of dialects (Spain, Colombia, and Chile). In fact, the results, according to contact with other varieties, revealed that Venezuelans had more exposure to the dialects tested in this study than Spaniards. These results might suggest that the geographical situation of the country plays a role in dialect recognition. Future research should explore patterns of immigration in Venezuela to examine possible explanations of the type of exposure that Venezuelans have. It seems possible to point out that this group of participants from Spain had less exposure to other dialects spoken in Latin America. Geographically, one can argue that Spain is more isolated in comparison to Venezuela with regard to other Spanish speaking areas, which means that Spain has less exposure to Latin-American varieties of Spanish. We could speculate that recent waves of immigration from Latin America to Spain might have an effect in groups of Spaniards who have had exposure to these dialects through contact with these new arrivals. Future research will have to explore this possibility. For the particular group participating in both of the experiments that we conducted, exposure to other dialects of Spanish was limited influencing the categorical perception of dialect variation.

There does not seem to be an effect for length of stimuli. However, listeners performed better in the passage task than in the sentence task. These findings indicate that more exposure to stimuli of a particular dialect might help participants in capturing indexical cues that might have an effect in improving rates of accuracy of dialect categorization.

Reaction time results correspond to the subjects' performance. Listeners needed less time to respond when they recognized a certain dialect. A clear pattern emerges from Spaniards in both the sentence and the paragraph experiment where listeners spent less time with stimuli from Spain. This is the variety that was more easily recognized by Spaniards and the one in which they took less time to select their answer. The same pattern is also observed for Venezuelan participants because they also spent less time in recognizing Spain in the sentence task as well as Spain once again and Colombia in the passage task. Colombia and Spain were easier to recognize for Venezuelan listeners. These results suggest that linguistic experience facilitates perception of indexical properties associated with geographical affiliation.

Those listeners with more contact with an area (e.g. family and friends) were more accurate identifying talkers from those areas. The findings are consistent across listeners from Spain and Venezuela as well as across tasks. These results have important implications for our analysis of the role of linguistic experience because they reveal that, in fact, more exposure to other varieties has an effect in dialect categorization.

There are mixed results for those who visited other countries. Results according to experience show better results for Spain than for Venezuela. These findings might reflect limited experience since the common factor was spending less than a month in the specific areas. In other words, the groups who participated in both experiments reported having limited experience with other varieties of Spanish. We might need to collect data from a more diverse population with more exposure to other dialects of the Spanish-speaking world to obtain more definite conclusions. Future studies will have to improve the definition of experience.

7. Conclusions

We now go back to the original research questions that guided the present investigation and present some suggestions for future research. The first question was related to whether naïve listeners are able to perceptually categorize speakers from different Spanish speaking countries. We found that listeners without formal training were able to recognize other varieties of Spanish with an average performance of at least 33% in the sentence task and of at least 35% in the passage task. Recall that for the sentence task Spaniards obtained 33% of accuracy, while Venezuelans obtained 35%. In the case of the passage task, Spaniards reached 35% of accuracy, whereas Venezuelans reached 44%. These results are consistent with findings from previous studies (Clopper and Pisoni, 2006; Clopper, Conrey, and Pisoni 2005; and Clopper and Pisoni, 2004a, 2004b, 2003-2004). In short, our findings show that naïve listeners of different Spanish dialect varieties can make judgments about an unfamiliar talker's country of origin without being trained on what to listen for. The ability to categorize not only their own dialect but also other Spanish varieties suggests that listeners retain in memory the varieties of their own language; not only on a local level but also across countries.

Our second question was concerned with what dialectal regions listeners are able to recognize more accurately. Our findings reveal that both Spaniards and Venezuelans were successful at identifying Spain with the highest rate of accuracy. Other areas that were identified include Argentina, Chile, and Colombia. Almost all the results in both experiments were above chance. Recall that Spaniards had a hard time perceiving stimuli from Colombia in the passage task.

Crucially important for the theoretical implications of our research is the role of linguistic experience in dialect categorization. The results revealed that contact (e.g. family and friends) has a positive effect on dialect recognition, while experience (e.g. travel) does not show a clear pattern for all groups. Regular contact with other dialects contributes to the encoding and retention in memory of information, which helps listeners to relate different speakers to where they are from. We have suggested before that experience has to be further investigated to obtain a clearer pattern because the sample participating in this research had limited experience with other areas of the Spanish-speaking world. In summary, a study with a more diverse sample is necessary to determine the role of experience, defined as traveling or living in a certain region. This group of speakers did not have much exposure to other dialect areas since they hardly lived in other countries, except for their country of origin.

The next question was concerned with whether length of exposure to the stimuli influences the correct categorization of the dialects. The results showed that participants performed better in the passage task than in the sentence task. This tendency is clearer for Venezuelan participants than for Spaniards since accuracy goes up for Venezuelans to 44% in the passage task. It might be the case that availability of more indexical features of geographical affiliation facilitates its perception.

This pilot study is the first step in improving data collection techniques for future analysis in different regions of the Spanish-speaking world. An examination of linguistic features is the necessary next step to understand their role in dialect identification. It is crucially important to determine what the linguistic features that facilitate geographical recognition of different dialects of Spanish are. Furthermore, an examination of dialect clustering by listeners is warranted. Listeners might not have selected the different regions randomly but as clusters of the major Spanish dialect areas distinguished by scholars like Canfield (1981), Lipski (1994); including varieties such as Peninsular Spanish, Andean Spanish, Caribbean Spanish, etc.

*Acknowledgments

This research was supported as a part of the *Speech Perception and Spoken Word Recognition* project NIH-NIDCD R01 DC000111. We want to thank David Pisoni, and Luis Hernandez for their help with the design of this experiment. We are grateful with Terrell Morgan for his collaboration in facilitating us the dialect samples used for the experiment. We also want to thank Pedro Diaz-Campos for his help in finding Venezuelan participants. We are also grateful for Jason Killam's and Elena Schoonmaker-Gates' comments and careful reading of an earlier version of this paper. All mistakes are our responsibility.

Appendices

Appendix 1: Stimuli

Read-aloud text used for creating stimuli

Hay más de trescientos millones de personas que hablan español. Principalmente, en España y Latinoamérica. Por razones históricas y geográficas, han divergido los varios dialectos de la lengua. No sólo existen diferentes acentos sino también diferentes léxicos. Se dice coche, piso y maíz en España; auto, apartamento y choclo en Chile; carro, departamento y elote en México. Sin embargo, las manifestaciones culturales del mundo hispanohablante arte, cine, deporte, literatura, música y televisión sirven para compensar la diversidad lingüística.

‘There are more than three hundred speakers of Spanish; they mainly lived in Spain and Latin-American. The dialects of the Spanish language have diverged for historical and geographical reasons. Not only there are different accents, but also different lexicons. One can say coche (‘cart’), apartment (‘floor’), and maiz (‘corn’) in Spain; auto, apartment (‘apartment’) and choclo (‘corn’) in Chile; carro (‘car’), departamento (‘apartment’) and elote (‘corn’) in Mexico. Nonetheless, the cultural expressions in the Spanish-Speaking world: art, film, sports, literature, music and TV, compensate for linguistic diversity.’

Appendix 2: Background Questionnaire

Cuestionario de Trasfondo Personal
Experimento sobre percepción dialectal
Indiana University

Sujeto número _____
IDmchoice/IDmchoicePP

Nombre _____

Profesión _____

País de origen _____

Edad _____

Nivel de estudios _____

Circula la respuesta adecuada:

- ¿Has visitado algún país hispano además de tu país de origen?
SI NO
- Si la respuesta fue afirmativa, ¿en qué países has estado y durante cuanto tiempo?

-
- ¿Has vivido mas de un mes en un país hispano además del tuyo? SI NO. Si tu respuesta fue afirmativa, ¿dónde?

-
- ¿Tienes algún familiar de otro país hispano diferente al tuyo? Si sí, menciona el país de procedencia.

-
- ¿Tienes amigos o conocidos de otros países hispanos? SI NO

- Si sí, ¿de dónde? _____

References

- Abercrombie, D. (1967). In *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Alfaraz, G. (2002). Miami Cuban Perceptions of Varieties of Spanish. In D. Long & D. Preston (Eds.), *Handbook of Perceptual Dialectology*, (pp. 1-11). Amsterdam: John Benjamins.
- Canfield, D. L., (1981). Spanish pronunciation in the Americas. Chicago : University of Chicago Press.
- Baker, W., Eddington, D., & Lindsey, N. (In Press). Dialect identification: The effects of region of origin and amount of experience.
- Boomershine, A. (2006). Perceiving and processing dialectal variation in Spanish: An exemplar theory approach. In Face, T. (Ed.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*. Massachusetts: Cascadia Press.
- Clopper, C., & Pisoni D. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change* 18, 193-221.
- Clopper, C., Conrey, B. & Pisoni D. 2005. Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology* 24, 182-206.
- Clopper, C., Pisoni, D. & De Jong, K. (2005). Acoustic characteristics of the vowel system of six regional varieties of American English. *Journal of the Acoustical Society of America* 118, 1661-1676.
- Clopper, C., & Pisoni, D. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects *Journal of Phonetics* 32, 111-140.
- Clopper, C., & Pisoni, D. (2004). Homebodies and army brats: some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change* 16, 31-48.
- Clopper, C., & Pisoni, D. (2003-2004). Some new experiments on perceptual categorization of dialect variation in American English: acoustic analysis and linguistic experience. *Research on Spoken Language Processing* 26. 29-46.
- Lass, N., Hughes, K., Bowyer, M., Waters, L., & Bourne, V. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America* 59, 675-678.
- Lispi, J. (1994). The classification of Latin American dialects. In *Latin American Spanish*. 3-33. London and New York: Longman.
- Milroy, J., & Milroy, L. (1985). *Authority in language: Investigating language prescription and standardization*. London: Routledge.
- Moreno Fernández, J., & Moreno Fernández, F. (1999). Madrid perceptions of regional varieties in Spain. In D. Long & D. Preston (Eds.), *Handbook of Perceptual Dialectology*, (pp. 295-320). Amsterdam: John Benjamins.
- Mullenix, J. W., Pisoni, D.B, & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.
- Nygaard, L.C., Sommers, M.S. & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.
- Pisoni, D.B. (1997). Some thoughts on “normalization” in speech perception. In K. Johnson & J.W. Mullenix (Eds.), *Talker variability in Speech Processing* (pp.9-32). San Diego: Academic Press.
- Pisoni, D.B. & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research*. Timonium, MD: York Press.
- Preston, D. (1993). Folk dialectology. In Preston, D. (Ed.), *American dialect research*, (pp. 333-377). Philadelphia: John Benjamins.
- Preston, D. (2002). The social interface in the perception and production of Japanese vowel devoicing: It’s not just your brain that’s connected to your ear. Paper presented at the 9th Biennial Rice University Symposium on Linguistics: Speech Perception in Context, Houston, TX, March 13-16.
- Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.
- Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties. The contribution of different linguistic levels. *Journal of Language and Social Psychology*. 8, 31-48.
- Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D. Long & D. Preston (Eds.), *Handbook of Perceptual Dialectology* (pp. 345-358). Amsterdam: John Benjamins.

Selected Proceedings of the 11th Hispanic Linguistics Symposium

edited by Joseph Collentine,
Maryellen García, Barbara Lafford,
and Francisco Marcos Marín

Cascadilla Proceedings Project Somerville, MA 2009

Copyright information

Selected Proceedings of the 11th Hispanic Linguistics Symposium
© 2009 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-432-4 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

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 and Inmaculada Navarro-Galisteo. 2009. Perceptual Categorization of Dialect Variation in Spanish. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, ed. Joseph Collentine et al., 179-195. Somerville, MA: Cascadilla Proceedings Project.

or:

Díaz-Campos, Manuel and Inmaculada Navarro-Galisteo. 2009. Perceptual Categorization of Dialect Variation in Spanish. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, ed. Joseph Collentine et al., 179-195. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2212.