ABSTRACT

KOHN, MARY ELIZABETH. Latino English in North Carolina: A Comparison of Emerging Communities. (Under the direction of Agnes Bolonyai.)

Previous studies of Latino English within North Carolina cite the opportunity to explore an emerging dialect as one of the prime motivations for their research (Carter 2004, Wolfram, Carter & Moriello 2004). And, while these studies identify community factors as creating difference between the varieties found in various locations (Wolfram, et al 2004), the larger context of Latino English must still be accounted for in order to understand how certain traits become field-site specific and which features prevail across the various emerging Latino communities in the Southeastern United States. This paper explores several linguistic variants within Latino English in two demographically distinct North Carolina communities to identify how the local ecolinguistic system and spatial setting affect an emerging dialect, and whether there are common traits between field sites that may become identified as regional features of Latino English.

The features to be studied include quotative frames, consonant cluster reduction, vowel space analysis, and Codeswitching (CS) patterns. This range of variables allows for multiple levels of comparison with previous studies of global English variants as well as with studies of interacting linguistic systems (Tagliamonte & Hudson 1999, Maguire, McMahon & Heggarty 2006, Blommaert 2003). Placing this information in the context of a cross-community analysis, this study demonstrates how structurally different spaces impact the distribution of linguistic variables, leading to community and regional norms with local and individual variation.

While initial analyses demonstrate community as well as idiolectal differences, looking at these patterns together demonstrates that local differences become established within regional norms. The effect of community difference leaves its imprint on the variety. Understanding linguistic diversity in this way demonstrates how multiple varieties come to exist in the same region: by interacting with the specific economic and demographic forces that shape each community.

Latino English in North Carolina: A Comparison of Emerging Communities

by
Mary Elizabeth Kohn

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

English

Raleigh, North Carolina
2008

APPROVED BY:

_____   _____
Dr. Walt Wolfram   Dr. Tonya Wolford

_____
Dr. Agnes Bolonyai
Chair of Advisory Committee

BIOGRAPHY

Mary Elizabeth Kohn graduated with a Bachelor's of Art in English and Spanish from Appalachian State University in 2002. Before returning to school she worked as a social worker, teacher, and interpreter. She plans to continue her education in the Fall of 2008 so that she can continue to learn how and why people talk like they talk.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1   Introduction: Space, Sociolinguistics, and the Study of Latino English

Understanding the nuanced interactions between space and social structures has been a goal of sociolinguistics since William Labov first asked directions to the "Fourth Floor" of various department stores in New York City (1966). The concept that shifting populations and changing social spaces influence language variation is similarly endemic to the discipline (e.g. Labov 1963, Wolfram 1974). So when a dramatic population shift occurs, it merits linguistic analyses in order to understand how languages evolve in contact.

The recent restructuring of Latino migration within the United States is one such watermark event as 50 percent of non-metro resettlement since the year 2000 now occurs outside of established regions such as the Southwest and significant increases are occurring in non-traditional metropolitan areas such as Raleigh-Durham, North Carolina, and Atlanta, Georgia (Kandel & Cromartie 2004). This fact has not escaped researchers who have begun significant acoustic analyses of the varieties emerging within these new communities (Carter 2004; Wolfram, Carter & Moriello 2004; Frazer 1996). Comparing such studies with previous work completed in California, New Mexico, and Texas allows researchers to identify common features and differences between the varieties that emerge, thus informing how ethnic minority dialects become established and to what extent they experience regional accommodation and variation (Wolfram, Carter & Moriello 2004; Thomas, Carter & Cogshall 2006; Carter 2007; Roeder 2006; Hartford 1978). Further, these studies cement Latino English's status as a structured dialect, rather than an inter-language variety as they demonstrate the constrained variation that occurs within the code, as well as identify the

relationship between learner varieties and Latino English spoken by native speakers of English who may or may not speak Spanish.[1]

While such fieldwork represents giant strides in understanding Latino English, difficulties for contextualizing the variety remain. Describing Latino community structures, while heavily theorized (Fought 2003, Santa Ana 1993), and while acknowledged as an important factor for interdialectal variation (Fought 1999, Gordon 2000, Wolfram, et al. 2004, Carter 2007), is rarely quantified. Such spatial structures frequently become an afterthought overshadowed by sociological variables that have more precedence for quantification. Part of the struggle to incorporate spatial context into the analysis of this dialect arises from the stated "complexity" of the Latino community (Fought 2003, Santa Ana 1993) where various levels of bilingualism, fluency, and cultural integration into American society stand out against the monolingual-normative ideologies present within American sociolinguistics and society as a whole.

To address these issues this study will 1) implement space as a crucial variable, and in the process: 2) contextualize Latino English in the Southeast within the broader processes of globalization, 3) create an analysis of linguistic features that is comparable to previous studies in both the Southeast and Southwest in order to understand how the spatial structures of communities influence the *acclimatization* of a variety, and 4) apply these notions to an understanding of how local and idiolectal patterns of variation interact with regional and global patterns. Further, the variables to be analyzed fit within current trends of global English studies allowing for broad comparisons with other regional, international, and ethnic minority varieties. These will include a full vowel space analysis for four individuals within

---

[1] See Sawyer (1959) for arguments supporting the theory of Latino English as an interlanguage variety and Fought (2003), Santa Ana (1993), and Wald (1984) for a sample of authors who oppose this view.

the community, plus a selective analysis of /æ/ and /o/ for a wider sample of participants, a consonant cluster reduction analysis, an analysis of quotative frames within each community, and an exploration of the use of codeswitching within each field site. The breadth of the variables under analysis allows for broader conclusions as trends identified in the data can be compared across various linguistic domains.

To lay the foundation for the linguistic analysis that follows, I will first explore how the incorporation of "space" into a theoretical model of analysis is not only useful but necessary for understanding language change that results from demographic shifts caused by the combination of economic and social factors commonly labeled as "globalization[2]." I will also describe economic and social factors that influence current settlement patterns within the Latino communities under investigation. Understanding how global restructuring affects this population in particular will aid in explaining the regional and local linguistic variation that occurs within the data analyzed in this study, and also why such data may diverge from findings of previous analyses.

1.1 Framing Space: The Intersection of Geography and Human Interaction

Connected with postmodernism, the reassertion of space in theory crosses disciplines from literary criticism to psychology and sociology. Such theories build on notions that social structures and space mutually interact and define each other (Soja 1989). This study relies on the work of these models to create a theoretical framework capable of analyzing the speech of immigrant populations which participate within globalization.

---

[2] Giddens (2002) identifies globalization as the flowing interactions of many global systems, including economic, cultural, political and scientific systems.

Relating the movement of populations and the subsequent shaping of communities to economic and other social forces, as the demographers referred to in the following section (1.2) do, sheds light on immigration patterns within the population under study. Such research reveals the ordered diversity within the Latino communities that are emerging in the Southeast, as well as motivations for resettlement. This information may explain regional inconsistencies where locations such as Siler City exhibit massive immigration, while other locations show only modest growth in the Latino population. Understanding such models allows for the prediction of linguistic patterns that emerge with certain types of community structures and settlement patterns. These trends will be discussed in greater depth in the following section, the Latino English Ecological Niche.

Spatial systems have also emerged as a theme within linguistics since 1972 when Haugen began to speak of "the ecology of language," which alludes to the relationship languages share with their environment. Two different schools of thought emerged from this theoretical framework: Mühlhäusler (1996) takes a more eco-centric approach by focusing on ecological language and resulting relationships with the natural world, but the more interesting approach for the purpose of this paper arises from Calvet (2006) and Mufwene (2001). These authors employ the ecological metaphor to explain the coexistence of multiple codes in the same society, as well as the growth and development of new codes out of both "mother" and "father" languages. Such a framework eases the work of variationists by examining how languages interact in the same environment, rather than attempting to keep languages separate as occurs in traditional genetic views of language change. In this model, the bilingual participant is no longer problematic within the context of a variation analysis. Rather than examining how one speaker uses his or her two languages, such a theoretical

4

framework identifies the "niche" of each language, whether a transplanted language can acclimatize (which refers to how a language adapts to a new environment in order to survive), or whether a language will acclimate (or survive through multiple generations) (Calvet 2006:117). Codeswitching, varying levels of fluency, and multilingualism all fit within the same environment, as they represent interacting codes within the same space. If these features are present they must be taken into account as they influence overall language trends within the community. As such, the question of whether to include non-native speakers of a language under investigation in a variation study becomes moot point[3], and the "complexity" of the Latino speech community becomes comprehensible.

Another facet of ecolinguistics is the acceptance of environmental impact on linguistic varieties (Mufwene 2001). This concept implies that geographical boundaries, whether political, economic or physical, influence linguistic variation and change. Linguistic analyses regarding the interaction of globalization and language change also touch on such concepts, as linguists like Blommaert (2003), Blommaert, Collins & Slembrouck (2005) and Coulmas (2005) explore how the economic, political, and technological effects of globalization, all of which change global topography, influence languages around the world. While these concepts largely rest in theoretical and bilingual studies, Britain (e.g. 1991, 2000), adopts human geography techniques when studying language change as he correlates shifting dialect boundaries with changing human interactions within the environment. As such, the theories can be quantified.

---

[3] Subject selection criteria will be explored further in the Participants and Interviews section (Section 2.2).

These approaches capture the nuanced spatial experience of populations undergoing relocation. This framework is particularly useful for immigrant populations as it is capable of following the fluctuating and multiple spatial experiences of migrants. It is also useful for analyzing multiple codes in the same community. Under this model, Latino English in North Carolina is far from an aberration, but an explicable variety that came into being just as prior varieties emerged from diaspora. This variety should be seen as the product of a changing linguistic environment where several languages, dialects, and social identities meet, interact, and seek equilibrium.

1.2 The Latino English Ecological Niche: National, Regional, and Local Trends

It is an uncontested fact that migration patterns for Latino populations shifted in the course of the last two decades from historic and frequently urban communities in the Southwest, Florida and New York to include more rural areas in the Southeast and Midwest (Durand, Massey, & Charvert 2000; Kandel & Comartie 2004; Kandel & Parrado 2005; Liaw & Frey 2007) (see figures 1 and 2). This becomes evident as the Latino population in North Carolina grew over 50 percent from the year 2000 to the year 2006 according to US Census estimates. While "chain migration" (Massey 1990) continues to promote immigration to the traditional and historically established Latino communities mentioned above (Liaw & Frey 2007) several factors influence the general dispersion of Latino populations across the US.
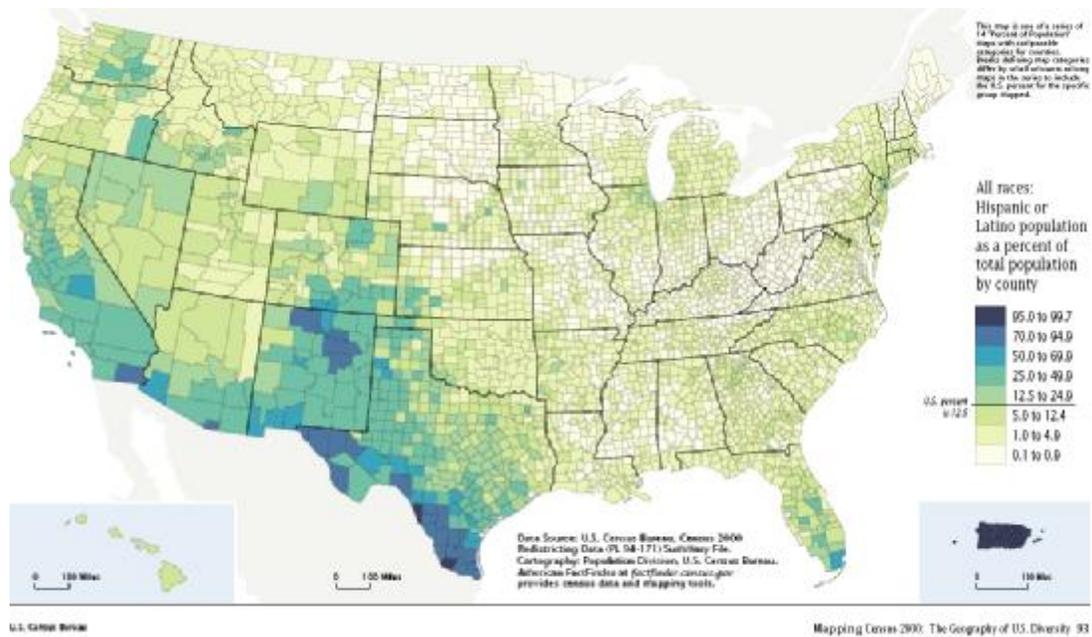
Figure 1. Population Density of Latinos in the US According to the 2000 US Census (Brewer & Suchan 2001:93)

Massey, Durand & Malone (2002) link these changes to legislation passed in 1986 (The Immigration Reform and Control Act) which both granted documentation to 2.3 million migrant workers while simultaneously tightening border security, thus unintentionally discouraging the return of many temporary and seasonal workers to their communities of origin. With legal documentation these workers experienced increased mobility and the opportunity to seek more permanent and better paying jobs: immigrant movement patterns within the United States began to align with other national population movement trends.

Changing immigration patterns also represent an extension of a change in economic structures. A crucial factor that influenced migration patterns consists of the economic restructuring that occurred over the past three decades which led to factories a) moving out of

urban centers and into areas that offer lower taxes, b) refiguring operations so that "unskilled" labor could be used[5], and c) recruiting migrant labor to fill the demand caused by high turnover due to poor wages, location, and safety considerations[6]. The meat processing industry in particular frequently engaged in this type of restructuring (Kandel & Parrado 2005). As such, it comes as no surprise that meat processing plants and textile plants are major employers of Latinos in rural areas (Kandel & Cromartie 2004). These new communities show significant differences from the traditional stable populations within California, New Mexico, Texas, Florida and New York in that growth is higher and communities are newer.[7]
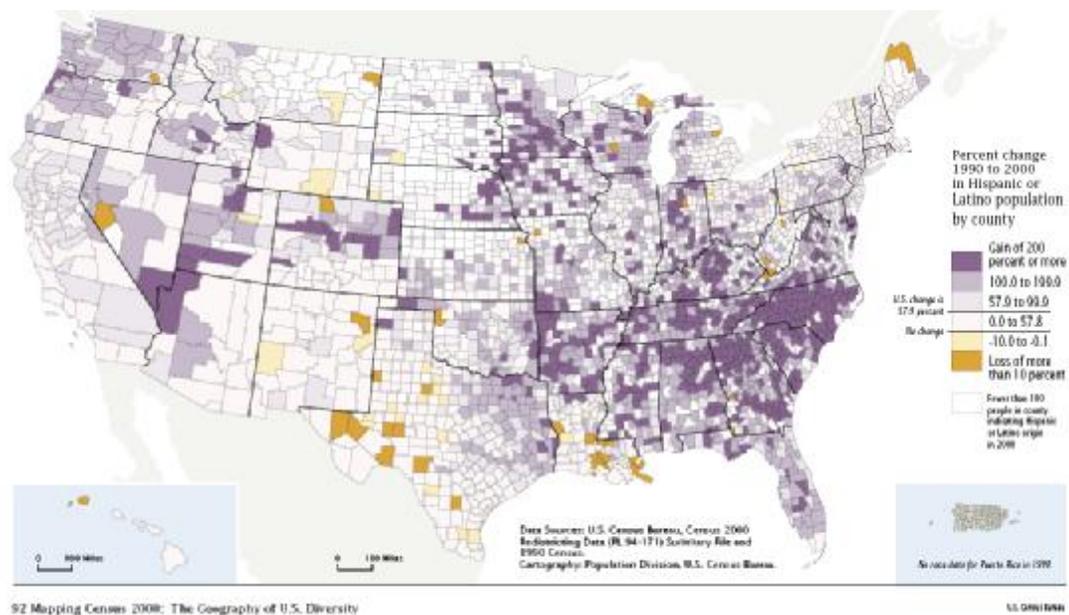
Figure 2. Population Growth for Latinos in the US Between 1990 and 2000 According to the US Census (Brewer & Suchan 2001:92)

---

[5] Although improving, Latino immigrants have significantly lower formal education levels than citizens of the US (Lowell & Suro 2002), thus making them more likely to fill "unskilled" labor positions.

[6] See Kandel and Parrado (2005) for a detailed list of studies supporting these assertions.

[7] In a recent report on the impact of Latino immigration in North Carolina Kasarda & Johnson (2006) found: "North Carolina's Hispanic population in 2004 totaled 600,913 or 7 percent of the state's total population, and accounted for 27.5 percent of the state's population growth from 1990 to 2004." Between 1995 and 2004, 38.2 percent of North Carolina's Hispanic newcomers immigrated from abroad, 40.2 percent migrated from another U.S. jurisdiction, and 21.6 percent were born in North Carolina. (Kasarda & Johnson 2006: xi)

8

While these patterns led to a more dispersed Latino population across the country, the amount of integration, monitored by living arrangements on a community and neighborhood basis, vary from location to location. For example, rural counties with high Latino population growth rates demonstrate increased insularity of the population when compared to counties with slower Latino population growth as these groups often live in specific neighborhoods in county seats while longer standing populations tend to move out of these towns and into more rural sectors of the county (Kandel & Newman 2004).

Understanding the economic and social motivations for migration promotes comprehension of the community differences between Latino populations within broad geographic regions such as the Southeast and the Southwest as well as in more specific locations such as Siler City, Raleigh, Durham, and Hickory, North Carolina. Because the Raleigh Durham Metro Region represents the twelfth fastest growing metropolitan community in the nation (Censusscope), Latino migration to the region mirrors other national migration trends (Massey, et al. 2002). As Raleigh and Durham represent high growth sectors promoting professional jobs and attracting large numbers of intra-national migrants to the region, one could speculate that Latino immigration to the area, particularly for those who are recent arrivals to the country, is due to the attraction of service industry and construction jobs, a factor listed as contributing to migration patterns in Liaw & Frey (2007). Partly due to lower education levels among Latinos when compared to other demographic groups in the US (Kohler & Lazarín 2007), it is likely that large percentages of Latinas are working in these less-skilled positions (De Anda 2005). A breakdown of predominating industries in the region shows the service sector to be the third largest employer in the region, after business

and healthcare, supporting this hypothesis (see table 1). The children in this study further

lend credence to this claim as they frequently mention that their families work at restaurants

or cleaning services, paint, do carpentry, or work in other such professions (see table 2).

Table 1. Raleigh-Durham-Chapel Hill MSA 2002 (North Carolina Profile 2002)
Predominant Industries

SIC Title	Employer Establishments	Workers Employed	Average Weekly Wage
Business Services	4,149	57,256	$ 774
Health Services	1,932	47,815	$ 742
Eating And Drinking Places	2,171	41,558	$ 254
Engineering & Management Services	3,033	29,457	$ 1,027
Special Trade Contractors	2,894	25,521	$ 590
Wholesale Trade-durable Goods	1,791	22,686	$ 1,160
Industrial Machinery And Equipment	154	17,022	$ 1,516
Food Stores	821	16,346	$ 332

Table 2. Professions of Participants and Participants' Family as Reported in Sociolinguistic Interviews[8]

	Factory	Service (restaurant and cleaning)	Construction/ skilled labor	Retail	Home-maker
Hickory	7	2	1	1	2
Durham	0	9	5	1	2

On the other hand, Hickory, North Carolina's attraction lies in the manufacturing jobs

present within the community. This too is reflected in the dataset as the majority of children

state that their parents work in various furniture and textile factories around the region (see

---

[8] Please note that some participants reported not knowing the nature of their parents' employment, while others did not discuss this topic during the course of their interview. These individuals are not included in this chart. Also, parents of sibling pairs who participated in the study are included only once in the chart.

table 2). Furniture and textile plants continue to be the number one employer in the region as of 2002 (see table 3), even as outsourcing continues to close factory doors in the region. Anecdotally, the tie between migration and factory jobs is supported as one older participant discussed how a factory recruited him during a lecture he attended as a university student in Santiago, Chile.

Table 3. Hickory-Morganton-Lenoir MSA Predominant Industries 2002 (North Carolina Profile 2002)

SIC Title	Employer Establishments	Workers Employed	Average Weekly Wage
Furniture And Fixtures	239	28,444	$ 557
Textile Mill Products	157	11,270	$ 530
Health Services	468	11,190	$ 638
Eating And Drinking Places	521	9,969	$ 228
Primary Metal Industries	19	5,453	$ 884
Business Services	431	5,119	$ 367
Wholesale Trade-nondurable Goods	194	4,296	$ 686
Food Stores	263	4,192	$ 263

Both areas differ from Siler City in that their larger sizes and tax bases allow for increased social services. Siler City's population represents the response of migration to changes in the meat processing industry as poultry processing plants are the number one employer in the area (see table 4).   Further, these employers are new, as the traditional economy consisted of agriculture (Downs, Fergus, Gagnon, George, Griffiths & Newman 2000), whereas Hickory's ties to textile and furniture production have lasted for over a century (Mohney & Phillips 1988). Because the furniture industry base in Hickory is a traditional employer within the community, while Siler City's food processing industries are relatively new, the former population may be more likely to work side-by-side with the

Anglo and African American populations that have been in the region for generations, while the latter are more likely to work with their ethnic cohorts. There is also a sizeable discrepancy in wages.

Table 4. Chatham County Predominant Industries 2002 (North Carolina Profile 2002)

SIC Title	Employer Establishments	Workers Employed	Average Weekly Wage
Food And Kindred Products	11	2,260	$ 421
Health Services	44	823	$ 436
Social Services	43	805	$ 406
Lumber And Wood Products, Except Furniture	27	707	$ 630
Eating And Drinking Places	53	679	$ 201
Special Trade Contractors	106	412	$ 501
Miscellaneous Retail	35	410	$ 554
Food Stores	41	402	$ 306

A few tentative conclusions can be reached regarding the Latino experience in each community. Within Durham, Latino populations working in the service sector and in construction are likely to earn substantially lower wages than populations attracted to the region due to the booming medical and business sectors. The Latino populations in Hickory, however, may be more able to achieve economic equity as the town's traditional populations are largely employed in the same factories where more recent immigrants also find employment. Siler City, on the other hand, is not likely to experience such integration as the employers in this city belong to an industry noted for its reliance on unskilled and often undocumented labor.

Understanding the interrelated economic trends not only allows the investigator to more fully understand the populations being studied, but also grants a predictive power to the

future. Cataloging the types of growth means that local differences within regions become explicable. Although the Latinos within these communities are part of the same migratory trends, the locations in which they settle offer different linguistic opportunities, as well as disparate likelihoods for integration. For the linguist this means that, with thorough analysis, ecolinguistic niches can be more fully understood, classified, and compared.

2   Field Sites, Participants and Methods

2.1 Field Sites

To understand how different environmental settings affect the linguistic homeostasis achieved within Latino English in North Carolina, this study analyzes the speech of two communities which, as discussed in the previous section, differ in economic and migratory patterns from both each other and from previous research sites that focus on Latino populations within the state (see figure 3).
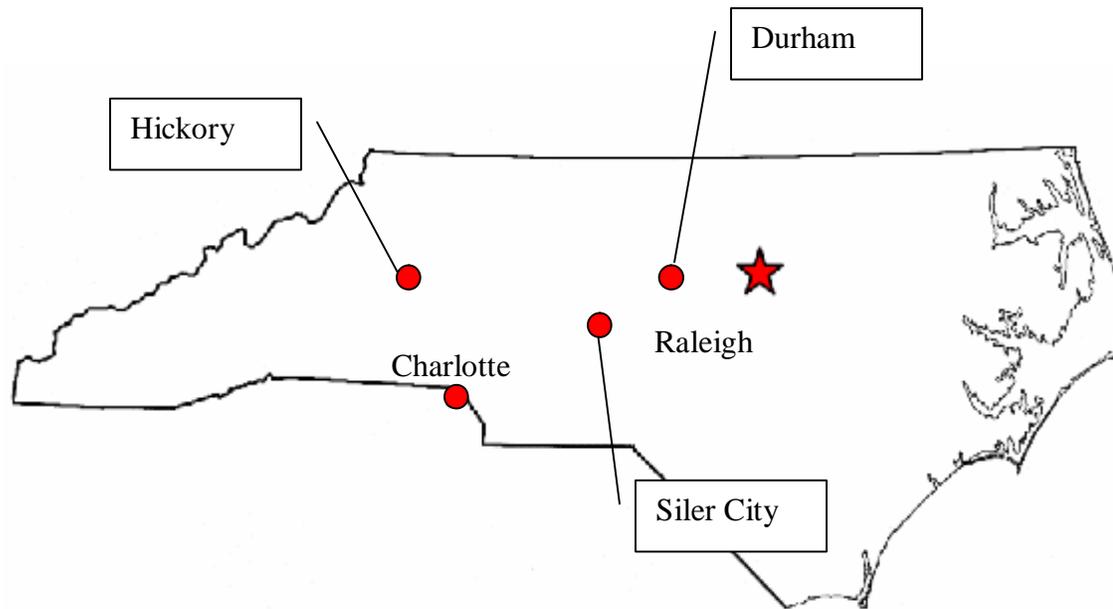
Figure 3. Field Sites (Netstate)

Hickory, North Carolina, is a mid-sized town located approximately an hour and a half northwest of Charlotte, North Carolina, and is overwhelmingly Anglo with a growing

Hispanic population of 7.7 percent according to the 2000 US Census (see figure 4). Durham,

North Carolina is a metropolitan region with an almost even divide between African

American and Anglo citizens, 44 percent and 46 percent respectively (see figure 4).

However, according to the US Census, the percentage of Latinos within the city is similar to

Hickory, 8.6 percent of the total population according to the 2000 US Census.
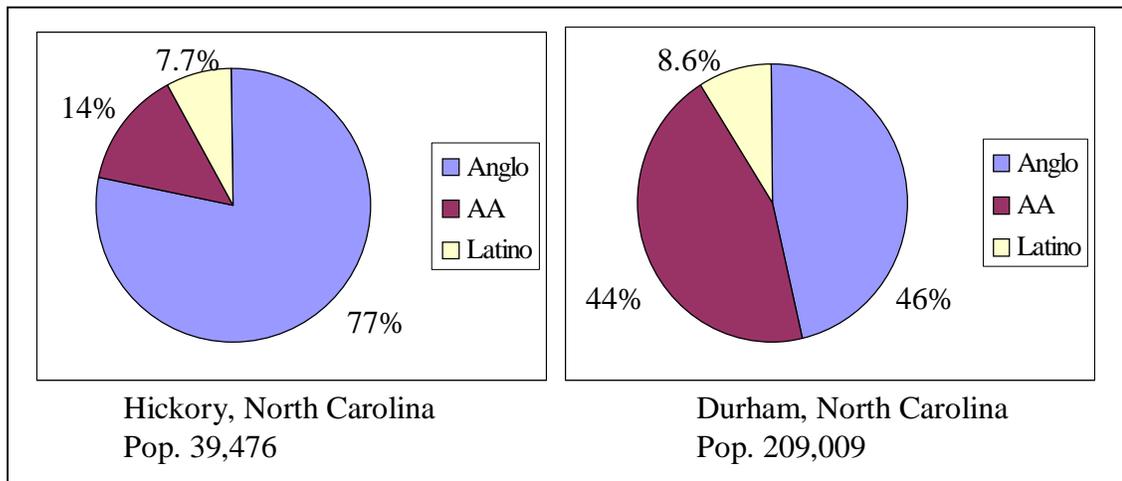
Figure 4. Demographic Data for Field Sites

2.2 Participants and Interviews

The data under exploration in this study was collected as part of the North Carolina

Language and Life Project and represent a sampling of the interviews taken from each field

site. A combination of methodological techniques was used to collect data, including

sociolinguistic interviews, social network analyses, participant observation, and informal

conversations with community members. Through these means significant ethnographic

information was collected over the course of a year and a half. Additional reading and

translation tests, as well as language proficiency evaluations were conducted with Durham

participants, although this information was only used to supplement vowel tokens for younger speakers within the vowel analysis.

Participants within Hickory, North Carolina, were contacted through "La Comunidad,"[9] a community support organization that offers adult ESL classes multiple times a week, home-work and childcare after school services for four days a week, legal advising, interpreting services, cultural activities, and community outreach. NCLLP staff collected all Comunidad data during the hours of the after school tutoring program that the center offers. All participants were selected either by La Comunidad staff or self-selected. Participants ranged from adult and peer tutor volunteers, staff, and those who come to La Comunidad for support.

Participants from the Durham field site represent student volunteers from two schools contacted through the English as a Second Language department who petitioned both current and former participants in the program, as well as friends of participants who did not participate in ESL. Interviews were conducted during school hours and teachers provided additional demographic and social observations.

To make the field sites comparable, only sociolinguistic interviews containing children between the ages of 8 and 16 were selected for analysis. This set of parameters has the added advantage of representing a wide range of length of residencies (LORs), from those who were born in the community to those who recently arrived. Any speaker who had lived in the United States for less than two years was excluded from the study due to difficulties in obtaining sufficient English speech samples.

---

[9] All proper names, including schools and organizations, are labeled with pseudonyms to protect the anonymity of the participants.

16

Although many studies of Latino English do not include participants with low proficiency levels, several considerations led to the exclusion of this criterion from participant selection. First and foremost, the research question for the study necessarily takes interest in the emerging nature of the community. Between 1995 and 2004, 38.2 percent of North Carolina's Hispanic newcomers immigrated from abroad, 40.2 percent migrated from another U.S. jurisdiction, and 21.6 percent were born in North Carolina (Kasarda & Johnson 2006:xi), necessarily indicating that large portions of the population are likely to be in the process of learning or acquiring English.

As over 78 percent of the community hails from locations outside of North Carolina, the question becomes how will each speech community reach an equilibrium within its new environment. There is precedent for studying new dialect formation among English-speaking populations in New Town communities where similarly-high levels of population shift occur (e.g. Trudgill 1986; Thomas 1997; Trudgill, Gordon, Lewis & Maclagan 2000; Mesthrie 1993; Kerswill & Trudgill 2005). However, the study of such communities as they become established within a contrasting host language remains rare (Mesthrie 1993)[10]. Ignoring such environments prevents a thorough understanding as to how dialect formation among communities associated with a heritage language occurs in the host language (see Carter 2004 for continued discussion of this theme).

This study takes the position that community differences may arise even within initial stages of community formation when the majority of speakers are still in the process of acquiring the majority language present in the broader context of the environment. Variation introduced by those acquiring or learning English, as their language is present in the speech

---

[10] Note that this study analyzed the koineization of the heritage language, not the majority language.

community, may germinate into community speech norms. Significant differences between communities within Length of Residency (LOR) status groups could indicate that local differences are emerging and may be maintained by future generations of English L1 speakers. While studies that excluded English L2 speakers were important in establishing the value of Latino English as a dialect that exists beyond imperfect language learning, ignoring these speakers now means ignoring the process through which codes travel to becoming a native speaker dialect, as well as the rich variation that currently exists in these communities. Because the main focus of this paper is the acclimatization of two speech communities, no speaker should be excluded on the basis of preferred code.

Twenty-one sociolinguistic interviews were collected with the 36 selected participants. As participants were often interviewed in pairs to reduce formality, this resulted in 16.5 hours of data. While most interviews ranged between thirty minutes to one hour, younger participants often gave shorter interviews.

These participants clustered into three LOR groups which will be used as an extra-linguistic variable to explore the effect of increased exposure to English on variation, as well as to identify whether certain types of variation is associated with those still in the process of learning English or with those who have fully acquired the language. The first LOR group identifies all individuals in the study who have lived in the United States for two to three years. These individuals generally participate in ESL classes. The individuals in the second LOR group have lived in the country for at least 4 years, but no more than 8. The final group identifies those participants who have lived at least nine years in the United States. Several of these participants were born in the U.S. and acquired English natively. All participants in this final group have had extended exposure to English through United States public schools.

Table 5. Participants (Age, Length of Residency)

Age	8-11	Lor	Age	12-13	Lor	Age	14-16	Lor	Age	Total
Hickory	1 male, 4 fem.			1 male, 2 fem.			3 males 3 fem.			14 total 9 females 5 males
LOR > 4 yrs	Dora Gail	3 3	10 10	Lou Catrina	3 3	12 12				3 females 1 male
4-7 yrs	Steven Jasmin	7 4	9 9				Marcela Charles Miguel	8 8 8	14 15 16	2 females 3 males
< 9 yrs	Emma	Life	9	Ivy	11	12	Darwin Sandra Robin	14 10 14	14 15 15	4 females 1 male
Durham	4 males 1 fem.			3 males 7 fem.			6 males 1 fem.			22 total 9 females 15 males
LOR > 3 yrs	Josh	3	11	Elizabeth Silvia Marisa Yolanda	3 3 3 3	12 12 12 13	Beatriz Cristobal	2 2	14 14	2 males 5 females
4-8 yrs	Marcela Jackson J. Y. J. V.	5 8 6 4	8 10 11 10	Jalinda Lori *	6 5[11]	12 12	Marcos Felipe Paco Tobi	8 7 6 5	14 15 15 16	7 males 3 females
< 9 yrs				Cleo Ronald Julio Jesus	10 life life 10	12 12 13 13	Lou	15	15	6 males 1 female

2.3 Areas of Analysis

The features to be studied include quotative frames, consonant cluster reduction, full vowel space for four participants along with a broadened analysis of specific vowels for a wider range of participants, and codeswitching (CS) patterns. This range of variables allows for multiple levels of comparison with previous studies of global English variants (Tagliamonte & Hudson 1999; Blommaert 2003), as well as with studies of interacting linguistic systems. Including an analysis of CS within the study will grant insight into how

---

[11] Although Lori was born in the US, she moved back to Mexico at age five and started school in Mexico. She moved back to the US two years later. As such, I place her in LOR group 2, even though she was born in the US

the heritage language of the migrant population interacts with systems already present within the ecolinguist space of the field sites.

One impetus for including such a wide range of disparate features, from phonological to syntactic, is to identify where community differences emerge. The paper will thus be ordered from the more syntactic variants, starting with CS, to more phonetic variants, concluding with a vowel analysis of select subjects. Identifying trends in community difference across varying linguistic domains suggests that community differences systemically impact the establishment of community norms.

3 Codeswitching across Field Sites

The impetus to include codeswitching (CS) in a variationist description of Latino English in North Carolina arises from two distinct, but converging arguments. From the ecolinguistic vantage point, in Mufwene's (2001) words, such an analysis can demonstrate the "ways in which a language may speciate in ecologies where it has been in contact with at least one other language" (19). Under this model, CS not only mediates language contact by promoting communication across speakers with varying levels of fluency in multiple codes, but also can lead to language change. Through CS, one code may eventually leave its imprint on another code through inclusion of lexical items, syntactic structures, or more subtle substrate effects like prosody, even if the heritage code disappears from use. From a variationist perspective, it is argued that CS represents a stylistic variable meriting analysis as CS becomes another symbolic linguistic resource within the bilingual repertoire (Bell 1984).

Combining these insights, locally-established CS norms become a site of variation across communities. Juxtaposing an analysis of CS patterns alongside other "monolingual" variables allows for a more complete understanding of how the two communities under analysis differ across linguistic modules. Yet, comparing such types of variation is rarely attempted (Angermeyer 2006). Within the context of a community comparison, differing community variation norms that cross multiple types of variables could reflect the effects of the local linguistic environment on the emerging variety in question, whether this be from internal sources, such as questions of identity, or external forces, such as increased contact with multiple codes in school and during social activities.

Approaches of exploring CS structural pattern variation quantitatively within and across communities are not new (e.g. Poplack 1980; Poplack, et al. 1987, Bentahila & Davies 1992; Treffers-Daller 1992, 1997; Cheshire & Gardner-Chloros 1998; Angermeyer 2006, Deuchar, Muysken & Wang 2007). Yet there is currently no analysis which compares pattern variation of CS and other types of "monolingual variation" for two separate communities with the same language pair. Such an analysis may inform the following questions: Although CS patterns are expected to vary across LOR, do they also vary by community even if the languages in question are the same? If such variation is present, does it exist in the structure of CS or in frequency only? What can CS tell us about the ecolinguistic environment of each community? Do these patterns share any similarities with other types of variation found within the community?

3.1 Codeswitching Patterns and Social Variation in Context

Our knowledge of the structural patterns of CS has grown over the past half-century as interest in CS increased (e.g. Muysken 2000, Backus 2001, Myers-Scotton 2002). Through these analyses, researchers have made general observations about the structures of switches that occur in speech. In order to quantify CS structural variation, a brief understanding of CS structure must first be reached (Myers-Scotton 1993b, Muysken 2000). At the most basic level, CS structural patterns may be divided into insertion and alternation, with the first category representing the combining of codes within or at the phrasal level, and the second representing the alternation of codes at larger syntactic boundaries such as clauses or sentences and/or speaker turns (e.g. Muysken 2000).

From a syntactic perspective, intra-clausal switches have received the majority of scholarly attention, as the question of what grammatical rules govern when and where switches occur motivates these analyses. However, the definition of what counts as a CS varies. A central debate on this topic is whether the insertion of a single word into a sentence of a contrasting code represents a CS or simply a type of borrowing. This type of structure is demonstrated in the following example: "I only know how to do *huevos* (eggs)," where the Spanish word for eggs occurs in an otherwise English sentence. Poplack (1980) considers single word switches as "nonce" borrowings and thus excludes them from her analyses of CS, while Myers-Scotton (1993a, 2002) considers borrowing and single word CS to be related processes on a continuum leading her to include single word switches in her studies of CS. While its status as a CS may be debated, both acknowledge that the practice of inserting lexical items from a contrasting language into a conversation is the most common way bilinguals alternated between codes.

There is also contention regarding CS structures that fall on the continuum between lexical insertion and sentential alternation, as the boundaries between whether such switches represent insertion or alternation within a clause can be fuzzy. In one of the first CS structural studies, Poplack (1980) reserves the term "intra-sentential" for multiple codes occurring in the same sentence, as in "Oh my god I don't know *cómo se llama,* (Oh my god I don't know *what it is called*)." Poplack states that this type of switch is: "a more complex or 'intimate' type, since a code-switched segment, and those around it, must conform to the underlying syntactic rules of two languages which bridge constituents and link them together grammatically" (1980: 589). Myers-Scotton (1993a) differentiates between a matrix language (ML), which determines the grammatical structure of a CS sentence, and the

embedded language island (EL), which is the contrasting language inserted into the matrix language. For example, in the following sentence two English ELs in the form of compound nouns occur in a Spanish matrix which determines the larger grammatical structure of the sentence: "Y los demás ya están en *high school* y *middle school*" (And all the rest are already in *high school* and *middle school*). Switching within a sentence can occur at multiple points, especially when languages are typologically similar, as demonstrated by this description of a traditional folklore figure: "And she have like *cadenas en sus pies* and *sus manos*" (And she have like *chains on her feet* and *her hands*)." Muysken (2000) uses the term "congruent lexicalization" for these frequent multiple intra-sentential switches. It is often difficult to determine whether this kind of switching represents multiple examples of insertion or alternation between phrases. Other types of intra-sentential switches may simply represent a change in code mid-sentence, such as between dependent and independent clauses, as in "She cried and she say, "*Aye mis hijos muertos*" (She cried and she say, *"Oh, my dead children*"). This type of switching may be considered alternation as it occurs at a larger syntactic boundary. As the boundary between alternation and insertion remains unclear, any quantitative study must confront this category of switching in order to design a replicable analysis.

Switches that occur at speaker turns are easier to identify. These clearly qualify as alternation switches, and avoid the grammatical issues that are attached to insertion switches, as shown in the following example:

> Beatriz: But she's not
> Cristobal: *Pero yo la digo sobrina (But I call her niece).*
> Beatriz: Okay, whatever.

Several studies have thus linked alternation CS to lower levels of proficiency (e.g. Bentahila & Davies 1992, Backus 2001, Poplack 1980, Poplack, Wheeler & Westwood 1987, Angermeyer 2006). Because they are easy to identify, it is of little surprise that this category is a common one for quantitative variation analysis of CS structure (e.g. Poplack 1988; Treffers-Daller 1992, 1997; Gardner-Chloros 1991; Bentahila & Davies 1992; Cheshire & Gardner-Chloros 1998; Deuchar, et al. 2007).

While CS grammars inform the researcher as to what switches are possible, researchers agree that actual CS behavior is further constrained by typological issues and social norms.[12] Quantitative comparisons of CS structure demonstrate that norms of usage not only depend upon proficiency, but also distance of typology for each language (Poplack, et al. 1987, Deuchar, et al. 2007). Barring external factors, communities that speak typologically similar languages, like Spanish and English, favor congruent lexicalization (Deuchar, et al. 2007). As this study analyzes CS between two typologically similar languages, this frequent intra-sentential type of CS may be expected. Previous studies of Spanish-English contact in the United States would support this expectation as well (Zentella 1997). All these studies lead to the belief that a comparative analysis of CS patterns is a legitimate way to identify community CS norms.

However, the setting, the identity of the participants, and the relationship between participants in the conversation also affect which types of CS structures are preferred. Previous comparative analyses indicate that the social and political climate of the region (Treffers-Daller 1992), interlocutor (Treffers-Daller 1997), style and speaker intention (Angermeyer 2006), and in some cases gender (Cheshire & Gardner-Chloros 1998) all can

---

[12] See Muysken (2000) ch.7 and Myers-Scotton (1993b) for an in-depth discussion of this theme.

influence the preference towards certain structural patterns of CS. Treffers-Daller (1997) found that, beyond speaker proficiency, the proficiency of the interlocutor also determined which type of CS structure was preferred in conversation. Further, as the political climate shifts, along with attitudes towards bilingualism, patterns in CS usage can shift as well (Treffers-Daller 1992). And, although studies of CS are largely restricted to in-group informal conversation, the impact of the data-collection methods, while assuring comparability, also is expected to impact the types of switches that occur (Myers-Scotton 1993b). Differences between the CS behavior in this study and that found in other Spanish-English communities, could relate to the interview setting, the political climate of the community in question, or proficiency. However, differences across the two communities within this study will not be due to data collection or proficiency, as these items are either controlled or quantified as factor groups. This will allow for a focus on whether quantifiable differences occur between typologically similar communities, pointing towards local variation. Because such an analysis will enhance understanding of how bilingual communities establish CS norms, this analysis attempts to compare CS structures across field site, LOR group, gender, and interlocutor pairing.

3.2 Methods

Tokens of codeswitches were taken from the main sociolinguistic interview analyzed for each participant. While it is a commonly accepted fact that CS occurs more often with in-group interactions among peers talking in an informal environment (e.g. Myers-Scotton 1993b), there is precedence for performing quantitative analyses on CS data collected in

more formal environments (Angermeyer 2006). As CS is connected with style, and as it occurs in multiple contexts of daily life, ignoring more formal settings such as sociolinguistic interviews prevents a deeper understanding of when and why CS occurs. Further, when performing a comparative analysis, controlling the mode of data collection becomes an important aspect of designing the study. Collecting data through a sociolinguistic interview at least partially controls for setting and interlocutor variability, as well as offers information regarding more formal styles of speech.

All of the sociolinguistic interviews were conducted by bilingual Anglo female graduate students who learned Spanish as a second language. The sociolinguistic interviews were framed in English, but participants were informed that the interviewers could understand and speak Spanish if the participant preferred. Although many of the participants also were included in Spanish interviews, this data was excluded from the analysis for consistency. Two interviews from the Durham field site that were conducted primarily in Spanish by the interviewer and primarily in English by the participants were excluded from analysis under this condition, although these interviews are analyzed for quotative usage and consonant cluster reduction. This left 33 participants for analysis. Of these 33 participants, all but 5 had at least one token of CS during the course of their interview. Lori and Rolando (who were both born in the United States) from Durham, North Carolina, were interviewed together. As Rolando's Spanish proficiency is limited, this likely accounts for the lack of CS for these two participants. Both Steven (LOR 4 yrs) and Catrina (LOR 3 yrs) from Hickory, North Carolina did not engage in CS behavior, as well as Jalinda (LOR 6 yrs) from Durham, NC, although no readily available explanation is present for these patterns. The remaining 16

individuals from Durham, NC and 12 from Hickory, NC had at least one token of CS in their interview.

Tokens excluded from analysis include proper names and place names. Names of cultural items such as food, dance, music, folklore figures such as *la llorona* and *la mano peluda*, and nicknames (e.g. 'My fat cousin we call him *gordo (fat)*') were included in the analysis. Codeswitches were bracketed according to speaker turn, so that each example of a codeswitch was analyzed within the period the speaker maintained the floor. Coding in this manner avoids the judgment calls necessary to distinguish syntactic boundaries such as sentence borders, which can be difficult to determine in naturally-occurring speech (Cheshire & Gardner-Chloros 1997). All examples of CS within the interviews analyzed were transcribed and assigned a code representing the pattern of CS observed in the turn. By these guidelines, 343 turns containing codeswitches were identified within the data. Immediately, a striking difference appears between the communities: 269 CS tokens came from the Durham data while the remaining 74 tokens were extracted from the Hickory data set. Each codeswitch was then categorized as a lexical and/or chunk switch, a multiple/phrasal/clausal switch, a turn-taking switch, or a block switch (see table 6 for examples).

Lexical and/or chunk switches describe all tokens of CS where a noun, verb, adjective, adverb, conjunction, preposition, interjection, minimal nominal phrase or compound noun was inserted into a sentence comprised of the contrasting language. The motivation for including compound nouns within a lexical insertion category arises from Backus's (2001) concept of semantic chunking, where these words form a unit of meaning and thus function like an individual lexical switch. Under these guidelines a switch such as "The other one is called *rueda de la fortuna* (Ferris Wheel)" acts as a lexical insertion

28

conveying one semantic concept of a particular amusement park ride. The fact that the speaker inserts this same compound noun into English sentences twice more in the interview indicates that the compound noun functions as a semantic unit. Only 11 such compound nouns occurred in the data.

Because each token was coded by speech turn, each CS turn was coded for either the presence or the absence of lexical/chunk insertions, although a general count of all lexical switches was made. The motivation for this coding technique is to make the counting techniques between categories consistent so that each category would be comparable for statistical analysis. As each speaker turn only provides one opportunity for a turn-taking switch to occur, while lexical switches may occur multiply within a sentence, these two categories represent differing and incomparable structural units: they do not have the same amount of opportunity for occurrence within an interview. Coding for the presence or absence of the type of switch per speaker turn makes the patterns of CS comparable in that they have equal opportunities to occur within each interview. This allows for comparison of favoring factors through a regression analysis. With this in mind, each category should be thought of as a description of a property of CS found within the speaker turns. Although 92 lexical/chunk turn tokens were documented this way, these tokens represent 151 words, some of which demonstrated multiple lexical insertions or compound nouns, as in "En Johnson[13] nada mas fui a *kindergarten* y *first* (In Johnson I only went to *kindergarten* and *first*)". In this example there are two English words inserted into the Spanish sentence, but instead of coding the turn twice, it is coded as one token of a turn where lexical insertion occurs. 61 of

---

[13] This is a pseudonym representing the proper name of an elementary school in the area.

these tokens were single word switches and 32 tokens were multiple examples of lexical

insertion within the same turn, minimal nominal phrases or compound nouns.

Table 6. Structural Categories for Switches

Category	Example

Lexical/ Chunk Insertion	Marisa: But I like the *Rueda de la Fortuna*
	          But I like the *Ferris Wheel*
	Christian: *Whatever*, tu tienes uno
	          Whatever, you have one

Phrasal/Clausal/Multiple insertion of varying sizes
 	
(sub-category of congruent lexicalization)
	Jackson: They were like, *'Que le muerda'* to the cake
	          They were like, '*Bite It'* to the cake
	Yolanda: Aye como que desea ser yo libra *because its like like beautiful* también *because I I like I love somebody* que es aires
	          Oh, how I want to be a Libra *because its like like beautiful* also *because I I like I love somebody* who is Aires

Turn-Taking Switch	Lou (D)- You know you call her that too. Julio-*Como? (what?)* Lou - you know you call her that too so why you trying to blame it on your brother
Block Switch	Cristobal: That's my niece Beatriz: *No es tu sobrina. ¿Cómo es tu sobrina si es hermana a mi- si mi mami es la hermana de tu papa? She is not your cousin. How is she your cousin if she is the sister of my- if my mom is the sister of your dad?*

The next category, phrasal/clausal/ multiple insertion of varying sizes, includes any turn

where more than one code are identifiable within a turn and where switches include larger

syntactic units such as the insertion of full phrases beyond minimal nominal phrases. This

category also includes CS turns where the speaker begins the turn in one language and then

completes the turn in another. As previously discussed, while lexical insertion is clearly a

form of insertion CS, and while turn-taking switches clearly fall under alternation CS,

phrasal/clausal/multiple switches prove more complex. The vast majority of the

---

[15] See Cukor-Avila (2002) for a partial literature review.

phrasal/clausal switches identified in the data are clearly alternation, with the speaker starting his or her turn in one language and ending in another or inserting a complete clause into a matrix sentence of a contrasting language. 66 out of the 83 phrasal/clausal/multiple switches fall into this category. However, 17 tokens appear to be examples of Muysken's (2000) congruent lexicalization, where due to the syntactic similarities between the languages, switching of various sizes occurs at multiple points within the turn (see example in figure 6). With such tokens it is unclear whether the switches represent alternation between grammars or insertion of an EL grammar into an ML frame. Because the latter represents a "deeper" type of CS (Myers-Scotton 1993b), these examples were identified and counted as a sub-category within the phrasal/clausal/multiple switch category.

Moving from insertion to alternation, all tokens of speech that contrast with the previous interlocutor's code but contained no CS within the token were categorized as a turn-taking switch. 150 tokens of turn-taking switches were extracted from the data. A final category of block switches was established to encompass switches that effectively renegotiated the language of conversation. Any turn switch that comprised an extended holding of the floor was counted under this category. If the speaker held the floor for three spoken sentences or more, it qualified as a block switch. The impetus for separating these tokens out from turn-taking switches comes from their discursive weight. A renegotiation of code represents a different type of act from a temporary switch as it is a marked choice (Myers-Scotton 1993a). There were 15 examples of block switches within the data.

Extra-linguistic categories coded for each CS speaker turn included gender, field site, LOR, and interlocutor (See table 7). The direction of the language switch was also coded for each token. Interlocutor status refers to whether the participant's speech was directed at the

interviewer, a peer, or to either and/or both the interviewer and a peer. Treffers-Daller

(1997) found interlocutor proficiency to correlate with certain types of CS behavior. Because

the interviewers are neither members of the communities in question, nor native Spanish

speakers, the types of switches directed at them may differ from those directed at peers. The

interlocutor was determined through discourse context.

Table 7. Extra-Linguistic CS Categories

Gender	Male	Female	
Field Site	Durham	Hickory	
Interlocutor	Peer/ community member	Interviewer	undeterminable
Preceding language	Spanish	English	CS
LOR	0-3 yrs	4-8 yrs	Over 8 yrs

3.3 Results
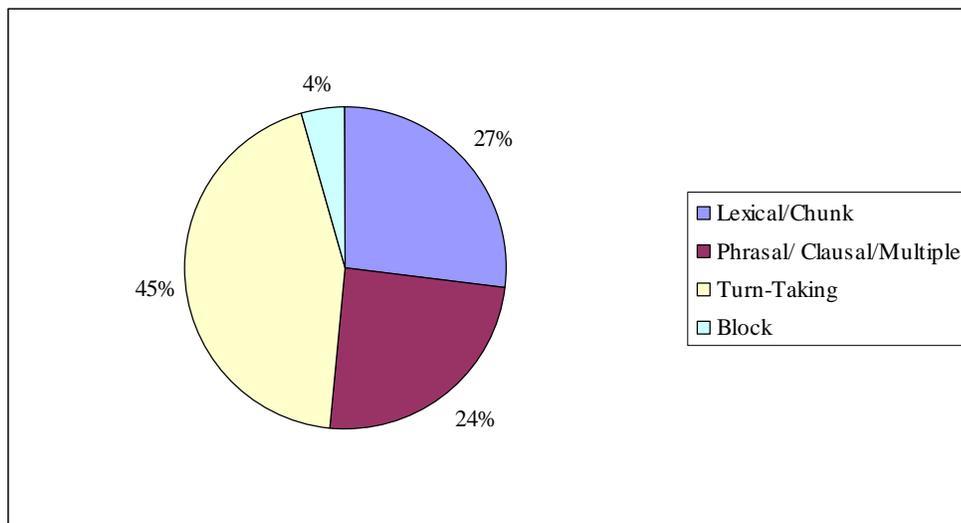
Figure 5. Total percentage of Switch Types

As shown in figure 5, turn-taking switches were the most common found in the data, at 45 percent of the total token count. Lexical/chunk insertion and phrasal/clausal/multiple switch tokens exhibited similar numbers of tokens at 27 percent and 24 percent respectively. Not surprisingly, block switches were the least common at 5 percent of the data. The frame of the English sociolinguistic interview likely suppressed extended code changes.

While the coding technique may partially explain the low numbers of lexical/chunk switches that occurred in the data set when compared with turn-taking switches (see 3.2), the lack of such switches must be influenced by other issues as well, as there were a total of 150 turn switches compared with the 144 words that were inserted into either an English or a Spanish frame. Part of the dearth of lexical/chunk switches may be attributed to the proficiency levels of participants who frequently codeswitched within interviews. Over 78 percent of the total switches, whether to English or to Spanish, were produced by individuals who had been in the United States for less than four years, as shown in figure 6. This left an equal number of tokens, 37, produced within the other two LOR groups. Such a large discrepancy likely indicates a shift in CS function from compensatory for those who have shorter LORs to strategic for those who have longer LORS, as described in Backus (1992). This also aligns with previous studies that link alternational switches with lower levels of proficiency (e.g. Bentahila & Davies 1992, Backus 2001, Poplack 1980, Poplack, Wheeler & Westwood 1987, Angermeyer 2006). It is also possible that insertional switches emerge as individuals become more aware of monolingual norms and ideologies that disfavor the use of multiple codes around out-group interlocutors.
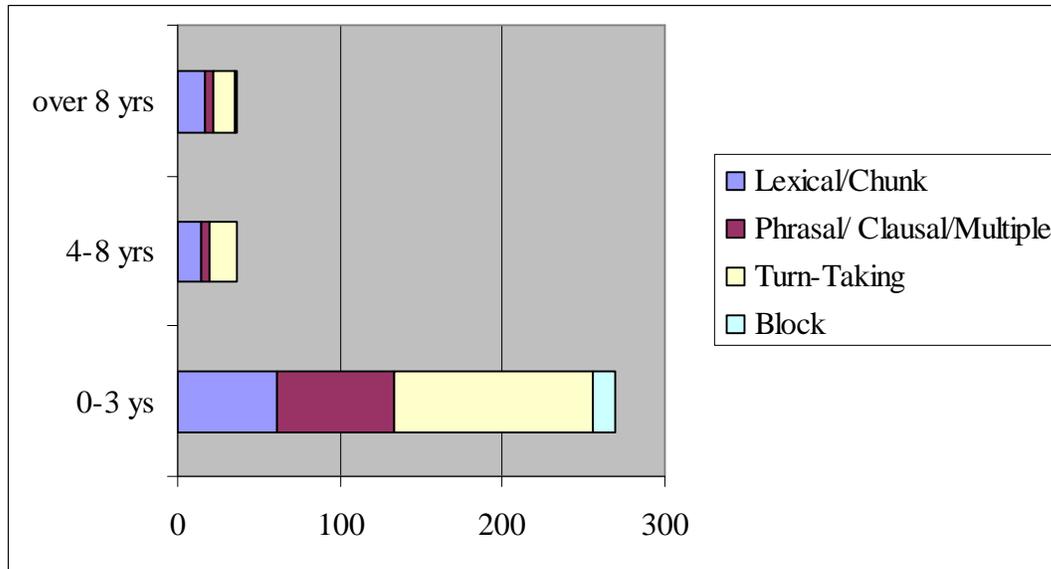
Figure 6. Number and Type of Codeswitches by LOR

As 79 percent of the CS tokens come from the Durham field site, CS appears to be more pervasive in this community. The percentage of lexical insertion switches in the Hickory data is slightly higher than in Durham (see figure 7), at 34 percent as compared to 25 percent. There is a larger discrepancy between the clausal/phrasal/multiple switches, with these tokens comprising 14 percent of the Hickory data and 27 percent of the Durham data. Notably, all 17 examples of the congruent lexicalization subcategory within the clausal/phrasal switches occurred in the Durham dataset. These differences would support the hypothesis that Durham is more accepting of CS in front of non-group members as intra-sentential switches tend to be associated with informal and non-standard speech, while single word norms violate prescriptive rules of language separation to a lesser extent. However, the difference in the distribution of CS structures between the two groups is not statistically significant (P = .093, Chi-square = 6.42).
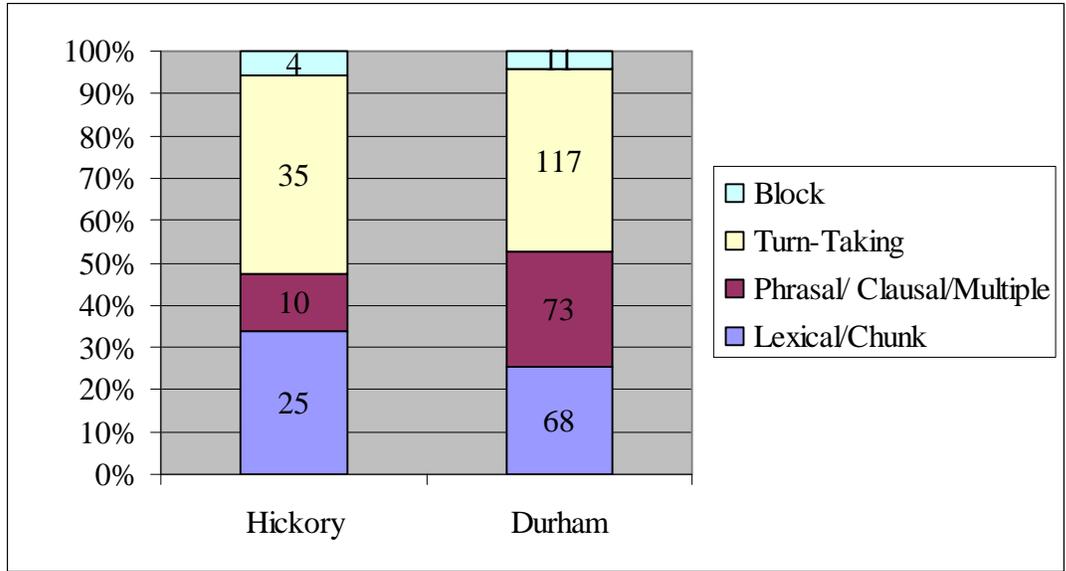
Figure 7. Comparison of CS Structures by Field Site

Remembering that 78 percent of the CS data came from participants who had been in the United States for only brief periods of time, the real difference between the communities lies in the CS behavior of the first LOR group, as becomes apparent in figure 8.
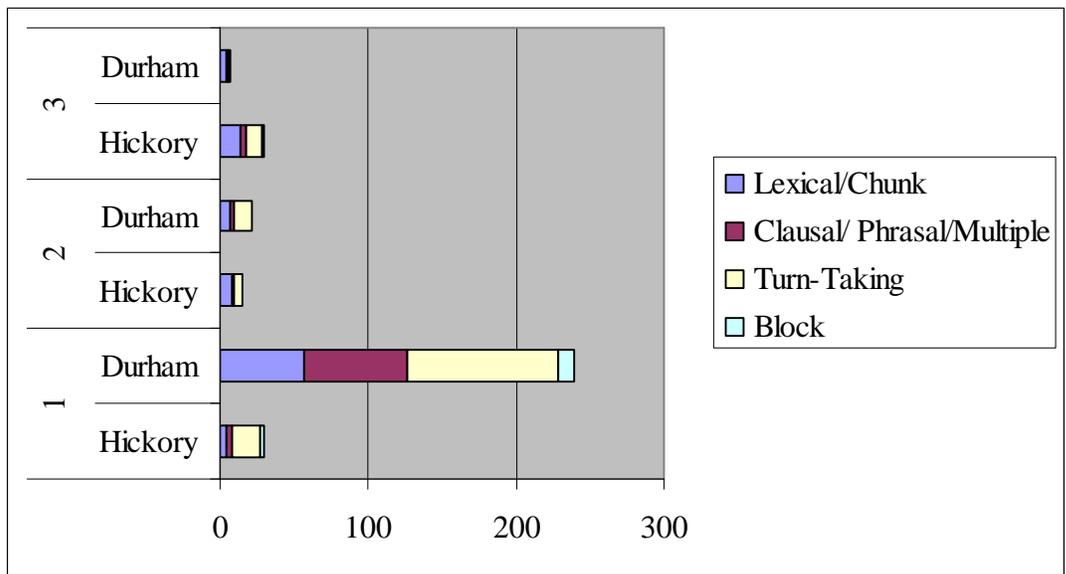
Figure 8. Total Token Count by Community and LOR Status

This difference is demonstrated by the fact that 6 of the 7 participants from LOR group 1 in Durham, NC engaged in at least 10 turns of CS within their interviews, while only one of the 4 individuals from LOR group 1 in Hickory codeswitched this frequently. Catrina (LOR 3 yrs) from Hickory did not have a single example of CS in her data. As CS behavior can be seen as "breaking the rules" (Cheshire & Gardner-Chloros 1997), this may reflect an avoidance of non-standard behavior among Hickory participants with short LORs. Alternatively, avoiding CS behavior may indicate a desire to conform to the native language of the interviewer, reserving Spanish and CS for in-group communication. More will be said about this topic in the discussion section.

To explore the interactions between structural type of CS data and LOR, field site, interlocutor, code of previous turn, and gender, a statistical model was built for the two most frequent types of switches – lexical/chunk insertion and turn-taking switches, using a step-up/step-down analysis in GoldVarb, a logistic regression modeling program designed for the analysis of linguistic variables. Such an analysis identifies favoring and disfavoring factors for the particular type of switch pattern out of the total CS turns found in the data. This clarification is important to note as it indicates that results do not inform how often such switches occur within conversation, but merely which forms are likely to occur given a set of factors when a CS turn does occur.

As shown in table 8, gender, interlocutor, and the code of the previous turn were all included in the best-fit model for turn-taking CS turns, but location and LOR were not. Factors that favor turn-taking switches include when the interlocutor is a peer, a preceding clausal/phrasal/multiple CS, and if the participant is a male. The strongly favoring effect of peer-interlocutor in the turn-taking category is of little surprise as CS behavior is known as

an in-group feature (Myers-Scotton 1993b). The favoring effect of phrasal/clausal/multiple

CS in the previous turn proves interesting. Such a switch in the previous turn may grant

licensure to the participant to freely choose a language, rather than compromise through

lexical insertion or another phrasal/clausal switch, as it indicates an acceptance of multiple

codes by the previous speaker.

That females disfavor turn-taking CS over other types of switches may be a result of

the data collection methods. As the interviewers were females, avoidance of turn-taking

switches may indicate a desire to use more accommodating styles of CS such as lexical

insertion. However, the effect of the interviewer's gender on CS behavior cannot be fully

known until comparable data sets collected by male interviewers are analyzed.

Table 8. Results of GoldVarb Run for Turn-Taking CS
Application Value = Turn-Taking   input = .445
Chi-square = 17.739   chi-square/cell = 1.0432
Log likelihood = -201.440

Factor Group	Factors	Factor Weight (Input and Weight)
Code of previous turn	English Spanish CS turn	.445 (.39) .478 (.44) .873 (.85)
Interlocutor	Interviewer Mixed Audience Peer/Community Member	.530 (.47) .333 (.29) .727 (.68)
Gender	Male Female	.601 (.55) .442 (.39)

One disadvantage of this statistical model is that the language of the switch remains

unknown. To fully understand the types of turn-taking switches that occur in the data, the

turn-taking switches were separated by language, site, and interlocutor in figure 9. A

statistically-significant difference emerges between the two groups (see figure 9). While the

Durham turn-taking switches occur in near-equal proportions regardless of interlocutor, the

Hickory participants clearly reserve Spanish turn-taking switches for in-group members. All

but five turn-taking switches in the Hickory data that were directed at the interpreters were

actually the participants responding in English to the interpreters' Spanish turn.
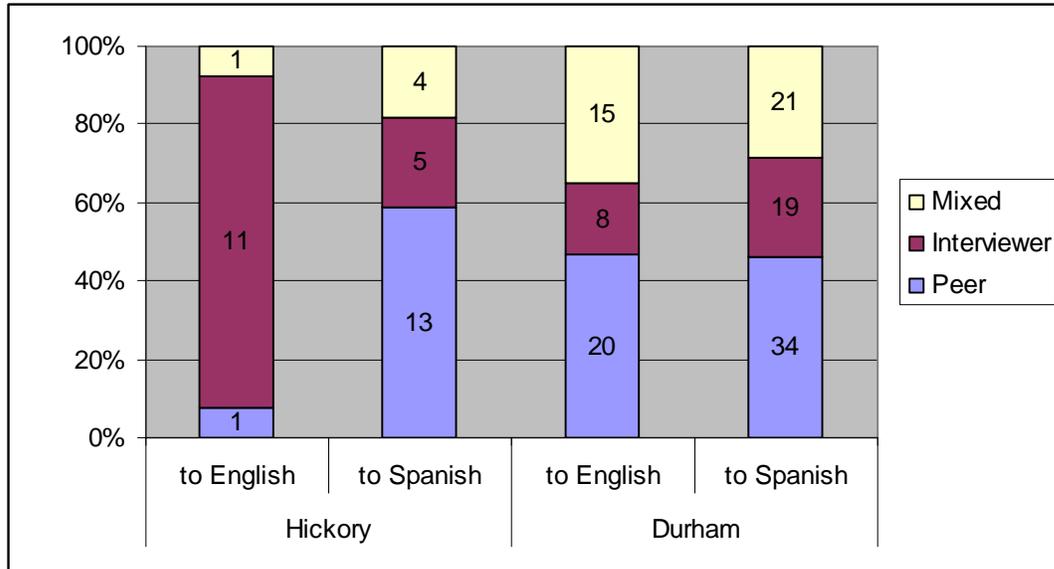
Figure 9. Person and Direction of CS by Community for Turn-Taking CS
The difference in the distribution between the groups is significant at p= .007 in an unpaired
t-test.

While shorter lengths of residency were not included in the "best fit" logistic

regression model for turn-taking switches, longer LORs appear to strongly favor lexical

insertion CS behavior within the "best fit" model for this CS structure (table 9). No good-fit

model could be obtained by including the previous turn's code as a factor, possibly due to the

fact that 82 percent of lexical switch tokens were Spanish words in an English matrix.

Again, the identity of the interlocutor is important as lexical insertion is favored when the

participant is speaking either directly to the interviewer or with mixed company. These

patterns encourage the hypothesis that lexical insertion may be seen as less of a violation of the structure of the English sociolinguistic interview so that those with higher levels of proficiency in English prefer this manner of CS.

Table 9. Results of GoldVarb Run for Lexical/Chunk CS
Application Value = Lexical/Chunk   input = .237
Chi-square = 7.6081   chi-square/cell = 0.8453
Log likelihood = -181.115

Factor Group	Factors	Factor Weight (Input and Weight)
Interlocutor	Interviewer	.603 (.35)
	Mixed Audience	.635 (.34)
	Peer/Community Member	.222 (.08)
LOR	1 (0-3 years)	.452 (.20)
	2 (4-8 years)	.631 (.35)
	3 (Over 8 years)	.700 (.42)

LOR and interlocutor do appear to impact the kinds of CS that occurs in the data according to these analyses. There appears to be community differences as well, but these mainly occur within the first LOR group. Further, patterns of turn-taking switches are significantly different between the two field sites. This answers many of the initial questions posed at the onset of this analysis. However, the intersection between LOR and field site has yet to be explored. Such an analysis may reveal more subtle factors in the structural patterning of CS.

Comparing the percentage of each type of CS structure by community and LOR status in figure 10 reveals that block switches do not occur in the Durham data for those who have been in the country over 3 years.   The block switches in LOR 3 for Hickory becomes more understandable when the individual tokens are analyzed. First, it must be recognized

that only 2 tokens of such switches were found in the data. Both occurred in David and

Lou's interview. Lou (LOR 3 yrs) stated a strong preference for Spanish and largely

refrained from using English throughout the interview. David frequently translated

questions, which led to block switches where Lou would ask David about these questions

before replying to David so that David could translate his answer back to the interviewer.
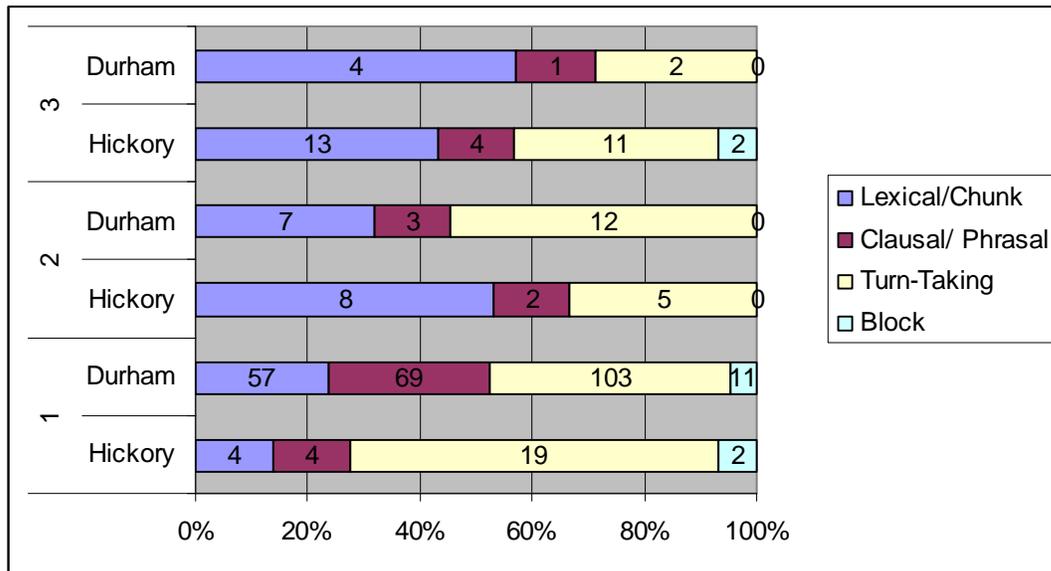
Figure 10. Percentage CS Structure by Community and LOR

Table 10. Statistical Comparison of Field Sites by LOR Group

LOR	Chi-Square	P
0-3 yrs (LOR 1)	**6.28**	**.043**
4-8 yrs (LOR 2)	1.89	.338
Over 8 yrs (LOR 3)	.305	.858

To identify whether or not the differences between the structures of CS used within

each community within LOR status groups, chi-square analyses were run excluding block

switches. The results are listed in table 10. The only difference that is found to be

significant is in the first LOR group. This raises interesting questions regarding the

convergence of CS variation over time. Do communities with similar language groups and proficiencies tend to use structurally-similar switches when in formal settings or around out-group members? Does the inclusion of a less-proficient speaker in the setting modify CS norms in such situations?

3.4 Discussion

To briefly summarize the findings up to this point, the amount of CS found in the data only significantly differs across community for those who have lived in the country for less than four years. This indicates that CS behavior is tied to competence. Turn-taking switches are favored over other types of switches when this switch is directed at a peer, possibly indicating that such switches signal a type of footing that demarcates side-conversations. It is also moderately favored if the switch is directed at the interviewer, but is disfavored if the switch is directed at a mixed audience. While Hickory participants tend to turn-switch to Spanish only when the interlocutor is a peer, Durham participants treat interlocutors in the same way, regardless of whether the interlocutor is the interviewer, a peer, or both. This may indicate that their CS behavior among the Durham participants is not a function of audience design, but more related to issues of proficiency.

Lexical insertion shows different patterns in that it is strongly favored by those who have been in the country for over 8 years while it is disfavored by those who have been in the country less than 4 years. It is favored when the interlocutor is the interviewer or is mixed company. Quantitatively, the difference between Hickory and Durham CS behavior is only statistically significant within the first LOR group, but Hickory participants show a tendency

to prefer lexical switches, while Durham participants show higher percentages of clausal/phrasal/multiple switches.

While the majority of these findings confirm previous research such as the importance of the identity of the interlocutor to CS structures used in communication (Treffers-Daller 1997), or the favoring effect of lower levels of proficiency and inter-sentential switches (Muysken 2000, Backus 2001), the community difference that appears among those with short LORs merits further discussion. It may be that Durham language learners are more likely to engage in CS behavior as a learner's communication strategy than Hickory participants. Such a strategy would be more likely if the majority of community members with whom they interact speak at least some Spanish. It may be that Hickory participants are less likely to rely on such a strategy as they have more contact with non-Spanish-speaking individuals. A survey of social contacts would be necessary to confirm such a hypothesis.

This pattern of community difference does not appear among the CS tokens for those participants with longer periods of exposure to English in either environment. Such a pattern may indicate a convergence of local dialect norms to a regional standard over time, or it may reflect the common practices of CS reached by populations with similar language repertoires interacting with similar interlocutors in the same speech setting (i.e. the sociolinguistic interview).

More subtle differences indicate a general trend toward either keeping the language separate or reserving Spanish as an in-group code in Hickory. This is demonstrated by the significant difference in turn-taking CS that occurs between Hickory and Durham, as well as the higher total percentage of lexically-inserted switches in Hickory. Whether this tendency towards

42

avoiding deeper CS in front of non-group members is an indication of increased contact with

monolingual speakers or an avoidance of a "non-standard" linguistic behavior is unclear.

The question that remains is whether or not such patterns align with other sociolinguistic

variants in these communities. Will there be a trend of difference found among the first LOR

group and similarity in the other groups for such features as quotative frames and consonant

cluster reduction? Such patterns may reveal how CS variation relates to other syntactic,

morphosyntactic, and phonetic variation found within the community, thus enlightening how

dialects acclimate in new communities.

4   Quotative Frames: Acquisition and Community Patterns

4.1 Background

Just as quotative *be like* has spread across the English-speaking world, so too have quotative frame studies.[15] Quotative frames consist of the syntactic bracketing of directly reported speech or inner dialogue,[16] as in, 'My brother *say*, "You so scaredy-cat, you should get out,"' and 'I *was like*, "no, no, no, no!"' Interest in the feature is not unwarranted as it represents an easily identifiable example of a global variant[17] (Tagliamonte and Hudson 1999, Buchstaller 2006, 2007), thus opening up possibilities to explore how local varieties interact with and pull from these rapidly spreading linguistic features (Blommaert 2003). Linked to movement associated with globalization, "geographic, social, and occupational" mobility encourage the spread of these forms (Chambers 2000:286), but the manner in which they are adopted by local communities is not always uniform.

Studies addressing global/local language change demonstrate that local speech communities establish their own norms of usage when acquiring global forms such as *be like* (Tagliamonte & Hudson 1999; Buchstaller 2006, 2007). Given the prominence of *be like* in such analyses makes the lack of ethnic minority studies on the topic that much more surprising (Ferrara & Bell 1995, Sanchez & Charity 1999, Cukor-Avila 2002, Singler 2001). If quotative frames can enlighten how local variation interacts with global change, then it should be possible to explore the interaction of ethnic minority varieties with majority community patterns as well. Such a study remains elusive, however, until diverse ethnic use of the form is more thoroughly documented.

---

[16] See Romaine & Lange (1991) for a discussion of reporting speech and thought in discourse.

[17] Chambers (2000) refers to these sweeping changes as indicative of "Global English," which he defines as "supra-national standard used by upper-middle class people everywhere" (285).

Further, while recent studies identify local norms for global variants, few have explored what "local" means. Do such structures exist on the state level, the city level, the community level, the group level, or even the idiolectal level? Also, although Ferrara & Bell (1995) comment that L2 speakers quickly acquire *be like* there is currently no direct analysis on how this process might occur. While quotative systems appear to be a topic that has been thoroughly exhausted, there is still much that can be learned through an analysis of quotative frames such as *say, be like*, and *go*. More importantly, as this feature is prone to establishing local nuances, quotative frames represent a syntactic variable that could be affected as a new dialect, especially with a heritage language, emerges.

This study will attempt to answer the following questions: What do the quotative frames of Latino youth in North Carolina look like and does length of residency impact these systems? Will the constraints of *be like* as used within the populations under investigation follow previously noted constraints within the United States? Will common patterns arise in quotative systems among these communities? How local is "local" when talking about quotative frames?

4.2 Previous Findings

While constraints on quotative usage do differ from country to country, certain constraints appear to be more common than others. The function of reporting thought[18] distinguishes *be like* from *say* in that the former can also occur with speech and mimetic sounds while the latter's function is largely restricted to reporting speech (Cukor-Avila 2002;

---

[18] Quotative *think* also is found to encode inner dialogue or thought, although its frequency varies across varieties (Tagliamonte & Hudson 1999).

Blyth, et al. 1990; Tagliamonte & Hudson 1999; Romaine & Lange 1991; Sanchez & Charity 1999; Buchstaller 2001). This division of quotative labor is cited as an explanation for *be like's* increased tendency to occur with continuous aspect or habituality, as constructed speech that recalls an actual speech event is punctilious in nature, while constructed speech that describes a state (such as a reoccurring thought) may be durative (Blyth, et al. 1990).

Even as content appears consistent, there is some disagreement regarding the constraints of grammatical subject for quotative frames. Outside of Ferrara & Bell's (1995) study, which found that *be like* was spreading into the third person, other United States studies agree that *be like* occurs more frequently in the first person due to its use to report inner thought or dialogue (e.g. Blyth, et al. 1990; Tagliamonte & Hudson 1999). However, this pattern may not hold true in African American English quotative systems, where Sanchez & Charity (1999) and Cukor-Avila (2002) indicate that first and third person are equally favored by the form.

The new quotative *be like* has also been linked to some important discourse functions within both spoken conversation and narrative. This frame enables speakers to balance face needs with footing during speech as it " . . . allows the speaker to retain the vividness of direct speech and thought while retaining the pragmatic force, but not the syntactic complexity, of the indirect mode" (Romaine & Lange 1991:264). Two resulting features of the frame include its use with hypothetical speech, as the semantic nature of the quotative relieves the speaker from committing to a verbatim repetition of the source quote (Buchstaller 2001); and its correlation with historical present, yet another device used to make narratives more "vivid" and "immediate" (Yule & Mathis 1992). This function could

result in the quotative's favoring interaction with the present tense; and several studies did find such a correlation (Blyth, et al. 1990; Yule & Mathis 1992).

Areas with less convergence include the favoring effect of gender,[19] distribution of other quotatives such as *go* and *think*, and distribution of the quotative totals across speech. Such differences have been attributed to the establishment of local norms within global constraints (e.g. Buchstaller 2007, Tagliamonte & Hudson 1999). But how distant must communities be in order to establish their own linguistic norms?

Blommaert (2003) theorizes that spatial differences cause irregularities in the transmission of global variants: "*Inequality, not uniformity, organizes the flows* and the particular nature of such flows across the 'globe'. Consequently, whenever sociolinguistic items travel across the globe, they travel across *structurally different* spaces, and will consequently be picked up differently in different places." Even as variants remain governed by important structural linguistic constraints, studying ethnic minority populations may demonstrate how structurally different spaces affect the acquisition of global variants, as such groups often experience space in a vastly different manner than majority populations[20] (Soja 1989). Because this study compares two different communities of Latino English speakers in cities that differ both in size and ethnic make-up, comparison of the groups may

---

[19] While Romaine & Lange (1991) and Sanchez & Charity (1999) found *be like* to be more common in the speech of women and Tagliamonte & Hudson (1999) found that British females preferred this form, Blyth, et al. (1990) found increased usage in males and Ferrara & Bell (1995) found that the effect of gender decreased across the four years of data they analyzed. Singler (2001) found that sex had no significant effect on *be like* usage. It is important to note that Singler did find elevated use of *be like* among female dyads, which may skew the effect of gender in this data, as all interviews were conducted by females.

[20] I am able to assert that the same location may be experienced differently by minority and majority populations due to the continued existence of de facto segregation where minority populations may be more likely to be placed in certain school districts or neighborhoods than those who live in majority communities. Specifically referring to the population under study, differences in nationality, immigration status, and legal documentation affect population mobility and access to government services.

indicate whether shared ethnicity and the political and social experiences that come with such an identity override the influence of the majority communities on language variation.

Such analyses of ethnic minority quotative patterns are currently sparse in the literature. While Ferrara and Bell (1995) included thirteen African Americans, twenty-eight Hispanics and three "Foreign students" in their study of the social distribution of *be like*, the analysis of ethnic minority quotative frames was confined to whether or not *be like* appeared in their speech with no further quantitative analysis of distribution or content. While only seven of the twenty-eight Hispanic participants evidenced use of *be like* in Ferrara & Bell's (1995) study, one of these seven was what Ferrara and Bell called a "level-four user" meaning that the participant exclusively used *be like* to frame all quotes. Outside of this study, Sanchez & Charity (1999) found that *be like* was the preferred quotative frame for Philadelphia African Americans. Cukor-Avila (2002) asserts that the form is hierarchically diffusing in African American speech from urban to rural centers while maintaining constraints found in other studies including the favoring effect of internalized speech, non-lexicalized sounds, and younger age groups. While *be like* patterned within AAVE (African American Vernacular English) in similar ways to the local Anglo varieties, she did find that the two groups differed in their use of *go* which was nearly absent from the African American participants' speech.

As Bayley & Santa Ana (2004) note, quantitative Latino English studies suffer from a dearth of syntactic variant analyses. Quotative frames are no exception, even as these authors anecdotally note widespread influence from Anglo systems among their speakers, with older Latinos demonstrating increased tokens of *go* and younger participants preferring *be like*. Fought (2003:107-109) similarly devotes two pages of her analysis of Chicano

English to discourse markers, *be like,* and *be all.* She concludes that most of her speakers use *be like* as a quotative frame, sometimes even exclusively, and regardless of class or gang status, which she also sees as an indication that ethnic minority varieties are permeable to majority influence. Thus, the literature currently contains five ethnic minority studies of quotative frames, only three of which address Latino English – and only in a cursory manner. This study will fill an obvious gap by providing a thorough description of what quotative frames Latino children in North Carolina use and how they use them. This description will provide the opportunity to explore the effect of LOR on quotative frame usage, as well as the chance to identify whether patterns are consistent over two different cities in the same state.

4.3 Methods

Following Cukor-Avila (2002), all instances of direct speech including dialogue and inner thought were transcribed from the data set. In addition, mimetic tokens were collected, although "clicks" (e.g. Fought 2003:79-80), and physical gestures were excluded as visual confirmation is necessary to differentiate such tokens from false starts.[21] Applying these conventions, 26 of the 36 participants used at least one example of direct quoted speech or thought within their interviews. A total of 403 tokens were taken from the data. However, as individual frequency of quotative use was widely variable (from no available tokens to 92 tokens), a 21 token cap was placed on each individual. Only six participants were affected by this cap, all of whom were over the age of 11. This group consisted of five females and

_____

21 However, as "clicking" is listed as a common feature of Chicano English (Fought 2003: 79-80), and as be like is often used to report such non-lexicalized sounds (Cukor-Avila 2002, Buchstaller 2001), future exploration of this topic with the appropriate recording methods could prove fruitful.

one male, with three females belonging to the Hickory field site, and the others belonging to the Durham field site. Because previous analyses of this data set indicate that such individuals skew site data all figures and numbers below reference the capped totals. 17 additional tokens were removed due to a) idiosyncratic structures specific to the idiolect of a participant (e.g. one participant demonstrated three tokens of *been+ verb* quotes as in 'They just *been yelling*, "Elizabeth, the bus is coming!"') b) status of direct speech vs. indirect speech is unclear (e.g. prosodic shifts indicate a change in footing, but the quote is subordinated with a clausal connector as in, 'She *said that*, "You got to um work hard too long."') c) there are insufficient tokens of a particular category for statistical analysis (e.g. there was only one token of a quotative frame with a second person subject). The resulting data set includes 290 tokens, over 169 of which are quotative *be like*.

This data was further coded according to gender, location of the field site, and the LOR of the participant. Three categories for LOR are consistently used throughout this paper (see section 2 for a table listing the participants according to categories, along with raw LOR figures): group 1 consisted of those who lived in the United States for less than 4 years; group two individuals had LORs of 4 to 8 years; and group three consisted of those who had been in the country for at least 9 years or were born in the United States.

Each token was coded for linguistic information as well (see table 11). This included the verb (or lack there of) used to bracket the quote (see table 12), the tense of the verb, subject, presence of AAVE features such as copula absence or invariant *be*, content of the quote, whether the quote was habitual in nature and its level of realis. The level of realis refers to whether the quoted item refers to an actual event or a hypothetical event (Buchstaller 2001).

Table 11. Linguistic Factor Groups for Quotative Analysis

Factor Groups	Factors	Example
Tense	Present	'*We're* just like, "Hey, I need to go to class."'
	Past	'My mom *was* like, "Calm down."'
	Future	'*I'll* be like, "Leave my Pippi Longstockings alone!"'
	Other	'*He'd* be like, "What?"'
Subject	First singular	'*I* was like, "No, but look, she's wearing a white dress."'
	First plural	'And *we* say, "yeah."'
	Third singular	'*She* said, "Whatchya doing?"'
	Third plural	'*They* be telling, "Naw, I used to bang this."'
	No subject	'And say, "What'd you do with it?"'
	"It" as subject	'*It's* like, "no."'
AAVE features	Copula absence	'He like, "Man, let me see your locker."'
	Invariant *be*	'They *be* like, "You go first," and all that.'
	No feature	'I was like, "Um, well, ok."'
Content	Speech	'My friend was telling me, "Tell him yes! Tell him yes!"'
	Thought	'I was like, "What is her name?"'
	Ambiguous	'She's like, "What?"'
	Mimetic	'He grabbed the cake and was like, "Boom!"'
Habituality	Yes	'They keep telling me, "Why you sitting with them?"'
	No	'I was like, "Hurry up, I want some food!"'
Realis	Realis	'My uncle, he be like, "The cake!"'
	Hypothetical	'Then my mom's gonna say, "You have free time but then you have to do this and this and this."'

Verbs coded were allowed to emerge organically from the data and the top five categories were coded individually, leaving only eight tokens outside of these categories. Examples of each form are listed in table 12. Tokens of *be like*, *say, zero, tell,* and *like (*with no *be*) were the most common forms, with *ask, go, say like, talk like, go like,* and *do* all exhibiting six tokens or less.

Table 12. Examples of Quotative Frames

Quotative	Example
Be Like	'I *was like*, "La la la la la," playing around.'
Say	'They *said*, "I wanna meet her"'
Zero	One participant describing what she would say if she saw *La Llorona*, a character from a traditional Mexican ghost story: "I'm gonna die. Oh God!"
Tell	'And Ms. B. *tell* her, "Yummy, come here!"'
Like	'*Like*, "Naw, that's boring."'
Verb + Like	'She *was saying like*, "Why are you all here?"'

Tense was divided into past tense, present tense, and future tense according to context. *Zero* frames and *like* frames, which have no overt verb, were coded as 'other,' as were conditional tokens and invariant *be* tokens. 'Other'-tense tokens thus refer to tokens that do not contain tense information due to the lack of a verb, or the presence of aspectual information such as invariant *be* or conditional constructions (e.g. 'Then, "No, I was doing my homework,"' 'They *be telling*, "Naw, I used to bang this "" or 'He'*d be like*, "What?"') Additionally, the presence or absence of the historical present was noted for each token, although this was not included under tense because the historical present refers to the combined effect of a past time reference and a present tense marking.

Following Tagliamonte and Hudson (1999) content of the quote was coded as direct speech, internal dialogue (referred to as "thought" in this study), and mimetic (non-lexicalized) sounds. To this list was added a category marking ambiguity, indicating that the content of the quote could be interpreted either as speech or as thought, as this pragmatic function is an important trait of *be like* (Romaine & Lange 1991, Buchstaller 2006). Romaine & Lange (1991) describe how *be like* affords the immediacy of a direct quote with the pragmatic distance of an indirect quote. Within this data set, several speakers forcefully yield this trait of *be like* to build up face during a story by assigning themselves a quote that

would be face-threatening to another individual within the narrative. Robin, a 15 year old female from Hickory, North Carolina, provides a clear example of this use of *be like* when describing a confrontation with a police officer. Robin relates how the police officer made fun of her mother after a car accident, incorrectly assuming that her mother did not speak English. She then tells how she filed a complaint with the police department, which, according to her story, made the officer feel bad. She concludes the narrative by relating the officer's apology and her subsequent response: 'I *was like*, "Don't cry."' Given the power disparity between the individuals, it is unlikely that such an exchange occurred; but, *be like* allows Robin to place this final evaluative remark in her story because the quote does not necessarily indicate direct speech. Ambiguity is more than a category for hard-to-categorize tokens, but is a functional trait of the frame that may cause it to be preferred over less discursively-flexible frames such as *say*.

In order to code for the discourse-pragmatic function of the quote, the surrounding context was taken into account. If the quote was part of a dialogue with a response or if the narrative indicated that the quote was spoken through the impact of the speech event on subsequent speech or actions within the narrative, then the token was coded as speech. However, if the quoted content does not elicit a response within the narrative, but is followed by continued dialogue, then the context indicates that the content was unuttered and the token is thus categorized as thought. Further, if no interlocutor is present within the narrative or conversation, then the token is assumed to indicate thought[22]. For example, one participant shares her feelings on having older siblings by saying: 'Sometimes I wish I had some older brothers or older sisters, but then I'm like, "Maybe I don't."' Her dialogue is clearly internal

---

[22] Although it is possible to speak out loud to one-self, this type of speech even simply reflects vocalized thought, as dialogue can only occur through interaction.

as she indicates no interlocutor. An additional signal that indicated the unspoken nature of a quote is the use of a third-person pronoun to refer to an individual present within the context of the narrative. For example, describing a conflict, one boy comments: 'I was gonna hit her back too but I was like, "Nah, she just gonna hit me back again."' Although the other individual was present in the context of the surrounding narrative, it is clear that the content of the quote was not spoken to her as the third person pronoun is used instead of a second person pronoun. While these signs allow for the identification of content as speech or thought, there are still ambiguous categories. These often occur at the end of dialogue sequences, as shown in Robin's example listed above. If no direct evidence from the surrounding context could be used to classify a token as speech or thought, then the token was classified as ambiguous.

Further, as 19 tokens combined quotative verbs with invariant *be* to produce such tokens as 'My mom *be like*, "Why you late?"' or 'She *be asking*, "What he talk about?"' the data was coded for presence of invariant *be* marking as well. However, while Mallinson and Kendall found frequent copula deletion paired with *be like* in their DC data set (personal communication), only 4 such tokens were present in this collection. Further, there was only one token of what Rickford, Wasow, Zwicky, and Buchstaller (2007) call *Intensive and quotative all* produced by a teenaged male with an LOR of less than three years living in Durham, North Carolina. This token was also excluded from analysis.

4.4 Results

Be like appears vibrant in the data set, comprising 58 percent of the data; although this still falls short of such studies as Sanchez & Charity (1999) who report that 67 percent of their quotatives for this age group are *be like,* and the 69 percent frequency within data from 2005 reported by Rickford, et al. (2007). The second largest quotative category is *say*, which comprises 21 percent of the data, followed by *zero*, *tell*, bare *like* (minus verb, but not copula deletion), verb +*like*, and other, all of which comprise less than 10 percent of the data (see figure 11). Although tokens of *think* were found within this dataset, this quotative was reserved for indirectly reported thought such as, "I was thinking that I came here to learn" or "they think they all that." This may indicate the importance of *be like* for reporting thought in dialogue as *think* does not serve this function within this data set.
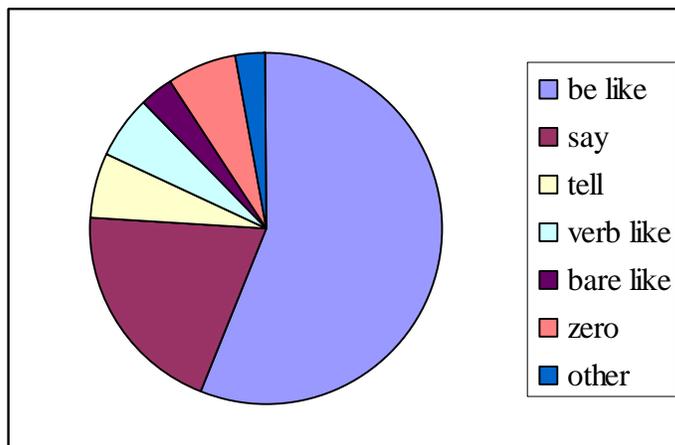

Figure 11. Quotatives Present in Total Data Set

When comparing the functions of the quotatives, this data set aligns with previous studies in that *say* is almost categorically confined to speech, while *be like* can frame speech, thought, and mimetic quotes. Notably, the majority of *be like* quotes do report speech,

demonstrating its vibrant incorporation into the quotative system within this data set.

Following the logic of the semantic meaning of *tell*, this quotative appears to function as *say*

and is also confined to the function of reporting speech. The *like-* derived forms of verb +

like and bare *like*, as in 'They gonna come and *say like,* "What you banging?"' 'They *say*

like, "Oh my god, your friend is so dumb,"' and '*like*, "Naw, that's boring"' appear to

demonstrate the discursive flexibility of *be like* in that they can occur in multiple categories

of content, although low token counts prevent definitive statements. Similarly, the majority

of *zero* tokens pair with speech, but there is evidence that this verb can also pair with thought

and mimetic tokens. Because several of the categories occur exclusively with speech, the

factor group of content will be excluded from the GoldVarb analysis that follows.[23]

Qualitatively, the populations under analysis utilize quotative *be like* and *say* similarly to

previous United States groups when comparing the content of the quotes framed by each verb

(e.g. Sanchez & Charity 1999; Ferrara & Bell 1995; Blyth, et al. 1990).

Table 13. Content of Quotative Frames

	speech	thought	ambiguous	mimetic	Totals
be like	121	15	29	4	169
say	60	0	1	0	61
zero	18	1	0	1	20
tell	17	0	0	0	17
verb like	6	0	0	1	7
bare like	6	0	1	2	9

A closer look at the function of *be like* across field site and LOR status group is

necessary in order to ensure that aggregation does not obscure more detailed group patterns

---

[23] See Singler 2001 for a discussion on why GoldVarb runs should not be used to analyze quotative frame content.

(see figure 12). While tokens of mimetic quotes framed by *be like* are scarce, the remaining

functions are all present across categories, and with no noticeable shift in distribution.

Speech is the most common quote content category across all groups, followed by ambiguous

quotes and then by thought. While tokens within the first LOR category are limited, the full

range of discourse functions for quotative *be like* is present by the second LOR group. No

noticeable differences between locations, beyond the lack of mimetic tokens for the Hickory

field site, are observable.
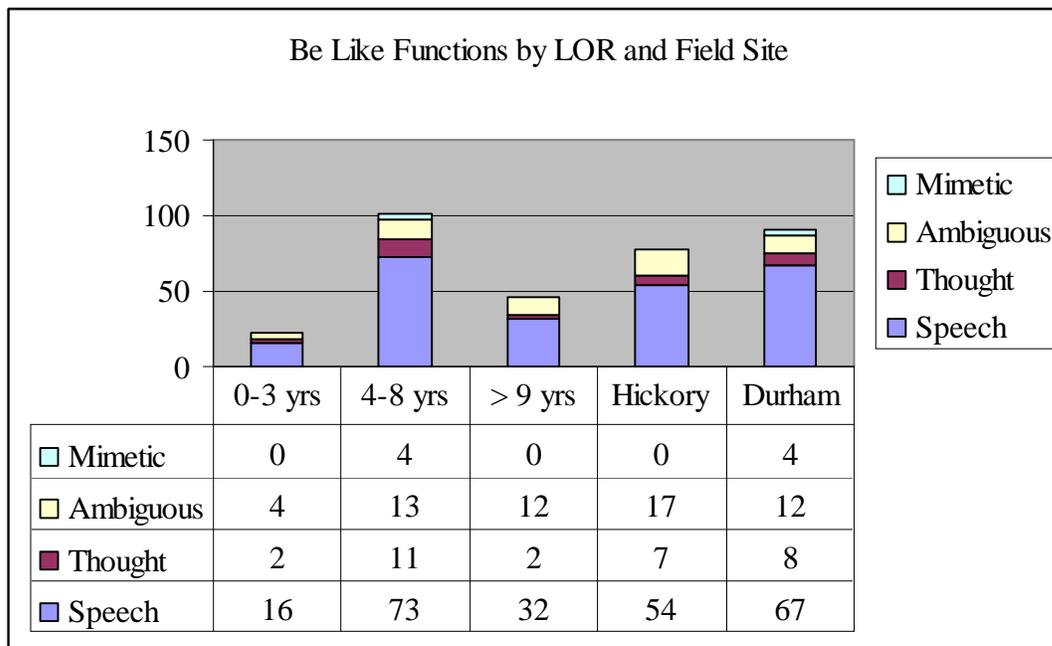
Figure 12. Be Like Functions by LOR and Field Site with Total Token Counts

	0-3 yrs	4-8 yrs	> 9 yrs	Hickory	Durham
Mimetic	0	4	0	0	4
Ambiguous	4	13	12	17	12
Thought	2	11	2	7	8
Speech	16	73	32	54	67

While content of the quotatives pattern so clearly that further analysis is rendered

unnecessary, important qualitative and quantitative differences in quotative usage exist

between both the field sites and the LOR categories (see figure 13). Perhaps the most

striking differences are observable in the diversity of quotatives observed in the Durham field

site and among participants with short LORs. Both Durham and LOR category 1 (0-3 yrs) demonstrate elevated use of the *say* quotative frame when compared to the other field site and LOR categories. The Durham participants also used two categories of quotatives (Verb + *Like* and *Tell*) that are absent from the speech of the Hickory participants.
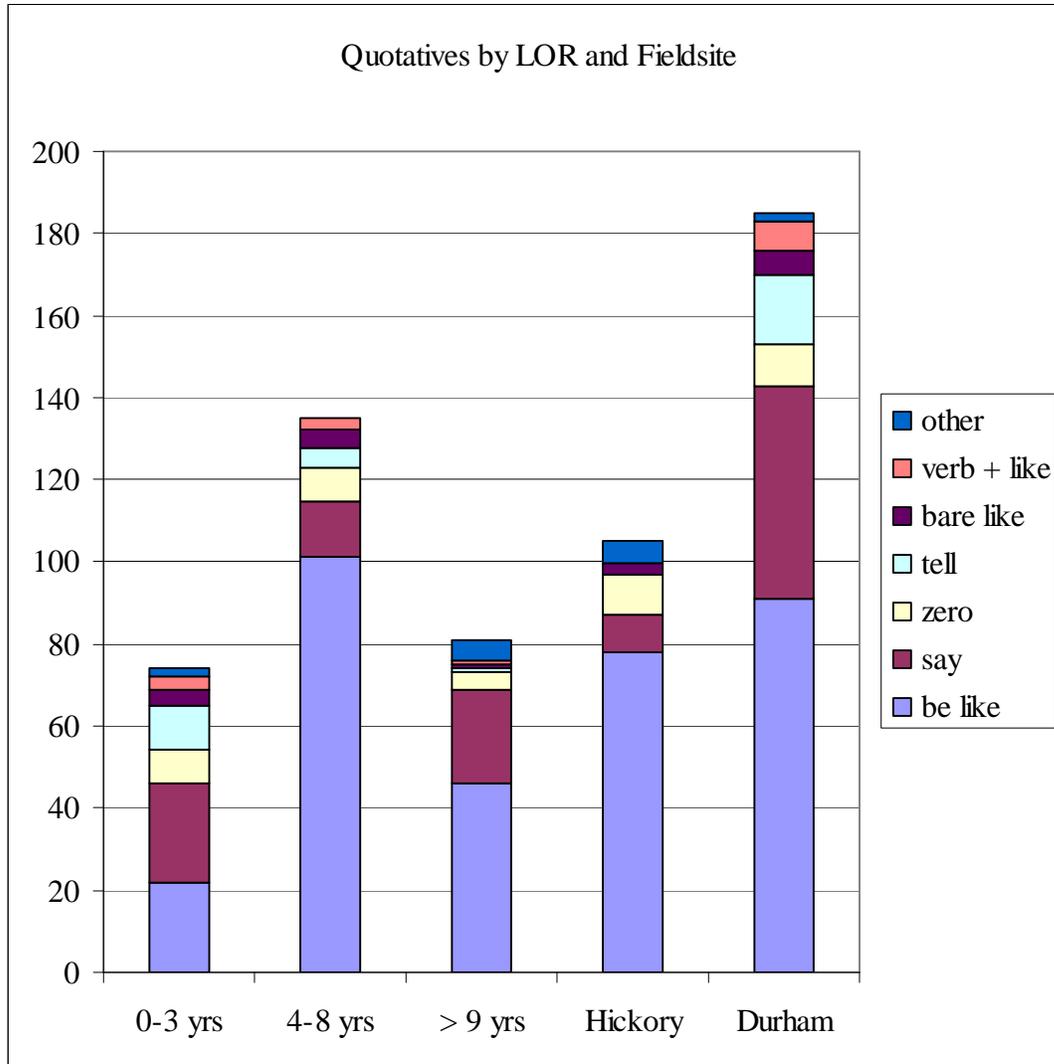
Figure 13. Quotatives by LOR and Field Site

It must be noted that this difference could be attributed to an imbalance in quotatives used by participants in the first LOR category across field sites as the Hickory field site only

contained four quotative tokens from one participant with an LOR of under 4 years, while the

Durham participants in the same category contributed 69 tokens to the data set. Although

there were other representatives of this LOR category from Hickory, two refrained from

reporting direct speech and the other switched into Spanish to do so. This may indicate a

difference in compensation strategies between field sites, but further data must be collected

and a separate analysis performed in order to test this hypothesis.
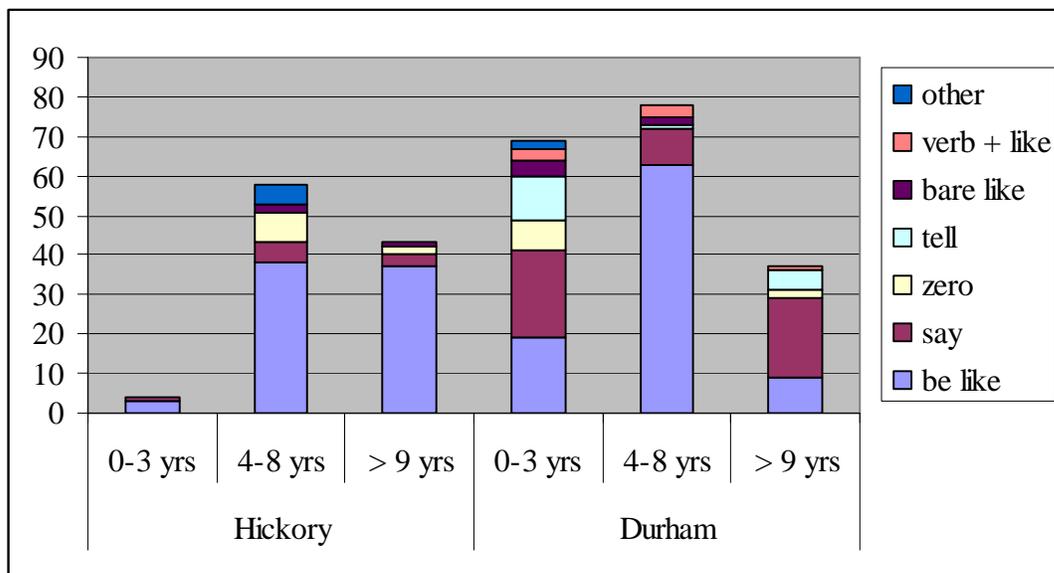
Figure 14. Cross-Comparison of Field Sites and LOR

However, the differences between the field sites cannot be solely attributed to

differences in quotative usage among participants with short LORs. When performing a

cross-comparison of field site and LOR qualitative and quantitative differences emerge (see

figure 14). The most apparent difference between field sites again is the use of the quotative

verb *tell* and verb + *like* in Durham. These tokens appear across LOR categories in Durham,

but do not occur in the Hickory data. Another surprising trend is the frequent use of *say*

among the third LOR category (LOR ≥9 yrs) for the Durham participants. A closer look reveals that of the four participants who fall in this category, one exclusively uses *be like* while another shows a preference for say, at 76 percent of his 21 tokens, showing widespread idiolectal differences within this group.

Several conclusions can be drawn from this analysis. The functions of the quotative frames appear to be consistent across field sites and LORs, and align with previous findings from research on the content of various quotative frames in that *be like* demonstrates a discursive flexibility that only *zero* can approximate as a quotative frame. Verbs semantically tied to speech acts, like *say* and *tell*, become restricted to the quotation of speech. Differences emerge between field sites as more quotative *say* tokens appear in the Durham data, and *tell* and verb + *like* exclusively occur among participants of this field site. This will be discussed in-depth in the discussion section.

A subsequent statistical analysis is used to explore the interactions of grammatical subject and tense on quotative *be like* and *say*. Each statistical model was built through a binomial step-up/step-down run including the factor groups of place, gender, LOR, quotative frame, subject, tense, habituality, and hypotheticality. The model with the "best fit" was then used to run a binomial one-level analysis. Table 14 demonstrates the results of a statistical analysis of the effect of grammatical subject, tense, LOR, and gender on *be like* using the statistical analysis package GoldVarb. The factor groups included in the "best fit" model are grammatical subject, tense, LOR, and gender.

Table 14. Results of GoldVarb Run for *Be Like* and Tense
Application Value = *Be Like* input = .611
Chi-square = 72.8305   chi-square/cell = 1.1205
Log likelihood = -147.868

Factor Group	Factors	Factor Weight (Input and Weight)
Grammatical Subject	First Person Singular	.652 (.75)
	First Person Plural	.616 (.72)
	Third Person Singular	.452 (.56)
	Third Person Plural	.245 (.34)
	"It"	.767 (.84)
	Zero	.562 (.67)
Tense	Past	.641 (.74)
	Present	.452 (.56)
	Future	.302 (.40)
	Other	.332 (.44)
LOR	1 (0-3 years)	.154 (.22)
	2 (4-8 years)	.741 (.82)
	3 (Over 9 years)	.448 (.56)
Gender	Male	.291 (.39)
	Female	.648 (.74)

Counter to Sanchez & Charity (1999) and Ferrara & Bell (1995), but following

Tagliamonte & Hudson (1999), Blyth, et al. (1990), and Cukor-Avila (2002), the first person

grammatical subject favors *be like*, while the third person subject disfavor it. "It" as a subject

also strongly correlates with *be like.* These findings are not surprising when taking into

account the previous analysis of content as the function of quoting thoughts promotes the

occurrence of *be like* in the first person. What is surprising is the favoring effect of past

tense with *be like* as studies that analyze quotatives and tense frequently report the favoring

effect of the present tense, which has been attributed to the use of historical present with *be

like* in narratives (e.g. Blyth, et al. 1990; Yule & Mathis 1992). Tense will be explored

further following the initial exploration of constraints on quotative *say*.

Other patterns that arise from the analysis include the favoring effect of the second LOR group (4-8 yrs). The other LOR groups (0-3 yrs and 9-life) disfavor *be like*, although the first group does so to a much greater extent. While gender appears to affect the model, no conclusions regarding overall gender patterning should be drawn from this data set due to data collection methods. As all interviews were conducted by females, and as female dyads have been found to favor the use of *be like* (Singler 2001), the true statistical impact of gender remains unknown.

Table 15 reports a similar run exploring the effects of grammatical subject, tense, LOR, and place as they occur with quotative *say*. A few adjustments were made to the model, including the inclusion of a dummy variable (Paolillo 2002) for the "other" tense category, as no tokens of invariant *be + say* or conditional *say* occurred within the data. [24] Further, no tokens of subject *it* occurred with say. The eight tokens of subject *it* were subsequently removed from the data for this run.

---

[24] No tokens of "Other" for *say* were available, although the categories are not inherently exclusive as *say* could occur with invariant *be* like *talk* does in the data, or could occur with conditional aspect. To place-hold this category, a dummy token was added.

Table 15. Results of GoldVarb Run for *Say* and Tense
Application Value = *Say*　　　input = .119
Chi-square = 85.3708　　　chi-square/cell = 1.4719
Log likelihood = -111.339

Factor Group	Factors	Factor Weight (Input and Weight)
Grammatical Subject	First Person Singular	.288 (.09)
	First Person Plural	.583 (.05)
	Third Person Singular	.691 (.18)
	Third Person Plural	.663 (.21)
	Zero	.739 (.28)
Tense	Past	.578 (.16)
	Present	.717 (.26)
	Future	.660 (.21)
	Other	
LOR	1 (0-3 years)	.543 (.14)
	2 (4-8 years)	.378 (.08)
	3 (Over 9 years)	.661 (.21)
Place	Hickory	.237 (.04)
	Durham	.655 (.20)

The design of this model may be less than ideal as *say* does not appear to favor

conditional use or use with invariant *be* while *be like* does. The resulting influence on the

model is the favoring effect of all proper (past, present and future) tenses. An alternative

model was constructed where all zero tense tokens were excluded from the data. This left

235 tokens for analysis. Table 16 demonstrates that the best fit model for this run included

grammatical subject, LOR, and place, but did not include tense. A similar model was built

for *be like*, but no "good fit" could be achieved.

Table 16. Alternative GoldVarb Run for *Say*
Application Value = *Say*   input = .193
Chi-square = 32.6071chi-square/cell = 1.4177
Log likelihood = -107.763

Factor Group	Factors	Factor Weight (Input and Weight)
Grammatical Subject	First Person Singular First Person Plural Third Person Singular Third Person Plural Zero	.268 (.08) .624 (.28) .630 (.29) .698 (.36) .829 (.54)
LOR	1 (0-3 years) 2 (4-8 years) 3 (Over 9 years)	.571 (.24) .338 (.11) .670 (.33)
Place	Hickory Durham	.260 (.08) .646 (.30)

Some tentative conclusions from these analyses include the observation that grammatical person does appear to correlate with particular quotative frames for these communities, with *say* being more common outside of the first person singular, and *be like* being favored in this environment. There does appear to be a difference in the amount of *say* quotatives used in each field site with Durham favoring the form and Hickory disfavoring the form. The effects of tense appear to be more problematic as no-tense tokens interfere with the construction of a statistical model that accounts for tense variation.

In order to explore the impact of tense further, tokens of *say* and *be like* were categorized according to present tense/ not present tense in figure 15 below.   A surprising pattern emerges where Durham rarely employs *be like* in the present tense, at under 20 percent of *be like* tokens (62 percent of 91 tokens occur in the past tense), while Hickory shows the reverse pattern, with 60 percent of *be like* tokens occurring in the present tense (33 percent of 78 tokens occur in the past tense). *Say* functions in the opposite way within the

Durham data. Over 60 percent of *say* tokens appear in the present tense and 35 percent in the past. The difference between field sites in their use of tense and *be like* is statistically significant at p < .001, but not so for *say*. This will be explored further in the discussion session.

Tense by Location
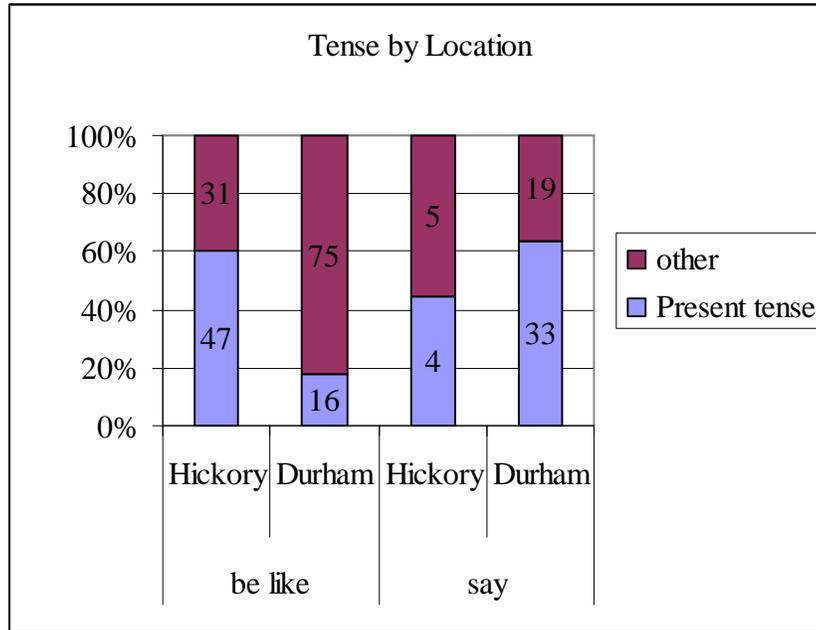
Figure 15. Percentage of Tense by Quotative Frame and Field Site
Difference in *be like* usage according to tense is significant at p < .001 with a chi-square of 32.7. Difference in *say* usage according to tense is not significant at p = .281 with a chi-square of 1.16

Remembering that the number of tokens for the first LOR group (0-3 yrs) is unequal between field sites, the data was broken down in a similar method and compared for tense usage in figure 16. Here, *be like* is shown to occur largely in the present tense among those with longer LORs (≥9 yrs) while it rarely occurs in the present tense among other groups. Again, this difference is significant at p=<.001. Mirroring this pattern, *say* occurs 87 percent of the time in the present tense in the first LOR group, but hovers just over 40 percent of the

time for the other two LOR groups (with 57 percent of tokens in the past tense for the second

LOR group and 47 percent for the third). This is significant at p=.003.
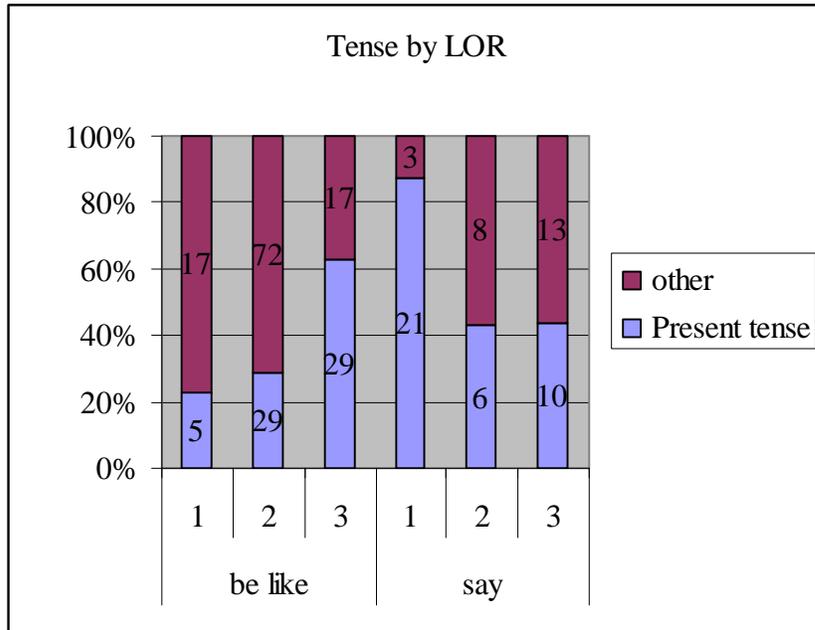
Figure 16. Percentage of Tense by Quotative Frame and LOR
Difference in *be like* usage according to LOR is significant at p< .001 with a chi-square of
18.2. Difference in *say* usage according to LOR is significant at p = .003 with a chi-square
of 11.9.

Moving from a discussion of tense to a discussion of AAVE invariant *be* and copula

absence, as presented in table 17, continued differences regarding field sites emerge. While

Hickory does have five examples of invariant *be + like*, as in 'They *be like*, "Let's go to this

dude's house,"' there are no examples of copula deletion combined with *be like* as in, 'He

like, "Man, let me see your locker."' Overall, copula deletion and invariant *be* tokens are

rare, comprising only 8 percent of the total data and the majority of these forms occur within

the Durham data. There is one innovative case from Durham combining invariant *be* with

tell to produce the token, 'They *be telling*, "Naw, I used to bang this."'

66

Table 17. AAVE-Marked Quotatives

	Hickory	Durham	0-3 yrs	4-8 yrs	≥ 9 yrs
Invariant *Be*	5	14	4 (Durham)	5 (Hickory) 13 (Durham)	1 (Durham) = *be telling*
Copula Deletion	0	4	0	4 (Durham)	0

4.5 Discussion

While the data analyzed from the two field sites correspond to previous studies in some ways, community-specific, LOR-specific, and idiolectal differences emerge. The data shows that the Latino youth who participated in this study do use quotative *be like* to frame thought, speech, and ambiguous quotes, regardless of field-site or LOR. Similarly, both *say* and *tell* are almost solely used to frame reported speech. This is demonstrated in all statistical GoldVarb runs which report a favoring effect of *be like* in the first person and a favoring effect for all other grammatical subjects for *say*.

However, an analysis of tense proves less clear-cut. Differences occur both across field sites and LOR groups for the use of present tense markings for *say* and *be like*. Specifically, *say* appears more frequently in the present tense among participants with short LORs, while the reverse pattern is shown for *be like*. Since the first studies of quotative *be like,* tense has been found to contribute significantly to patterning of the quotative (Blyth, et al. 1990), a trait that was tied to *be like's* correlation with the historical present in narrative, as both function to make narratives more "immediate" and "vivid" (Romaine & Lange 1991, Yule & Mathis 1992). If both quotatives were found to have few present tense markings in short LOR groups, but more frequent markings in longer LOR groups, a hypothesis regarding

the acquisition of the historical present could be built. However, *say* and *be like* pattern very differently within the data.

A stronger hypothesis draws from second-language acquisition studies. As *be* is an irregular verb that has high frequency, those learning the language would tend to mark the past tense more consistently with this verb (Doughty 1991). Conflated with the difficulties of acquiring the historical present, which is a sophisticated discursive strategy whose function may be difficult to observe for a non-native speaker, the historical present may emerge late within those still acquiring English (Wolfson 1982). This would lead to more frequent past-tense marking for quotative *be like*. Such a hypothesis would also support the dramatic rise in the use of the present tense for those with the longest LORs as they have spent the majority of their lives speaking English and thus would be more likely to incorporate the historical present into their speech in a native-like fashion. Further, *be like* has been shown to accept aspectual marking in this data set, including through invariant *be* constructions and conditional statements. This would also reduce the number of tokens in the present tense. 22 tokens of the 169 *be like* tokens fall into this category.

Say, on the other hand, suffers from the phonological vulnerability of its past tense construction. As all past-tense *say* forms are marked with the morpheme –d, the past tense morpheme can easily be deleted without forfeiting semantic meaning. The high number of present tense tokens for *say* among the first LOR group likely reflects tense unmarking, rather than historical present.

Another interesting observance is the increased use of *be like* and past tense marking for *be* in the second LOR group. This could represent a transitional period toward more

native-like English as these speakers may be keenly aware of the need to mark tense, but not yet familiar with the historical present. As these participants are likely to have experienced inclusion in ESL classrooms, as well as yearly proficiency testing, their sensitivity toward standard norms may be elevated, A closer look at tokens with past time reference but present tense marking across LOR groups is necessary to confirm such a hypothesis.

Yet length of residency cannot explain all community differences observed within the data. One such difference is the use of *tell* and *verb + like* among Durham participants, something that is categorically absent from the Hickory participants' speech. Further, *say* occurs more frequently in the Durham data, as well as invariant *be* and copula deletion. Compounded, these differences demonstrate community-level divergence thus indicating that the varieties are *acclimating* in different ways.

As with other varieties, quotatives used by Latinos in North Carolina pattern consistently with global norms in some respects, such as content of the quotatives. However, the inclusion of non-native speakers in the study demonstrates how features such as tense are affected by language learning trends and strategies. Although effects such as tense unmarking and lack of historical present appear to dissipate among participants with longer LORs, the heritage of language learning may continue to influence the variety in more subtle ways. Durham's innovative uses of *tell* and *verb + like,* as well as high token counts for *say,* could possibly be attributed to the influence of L2 English speakers on the variety, as a Spanish equivalent to *be like* is not readily present, but is so for *say.* A comparative study with Anglos and African Americans, as well as direct speech discourse strategies in the Spanish of the participants in the community would be worth pursuing in order to identify the source of these community differences.

5   Consonant Cluster Reduction

Consonant cluster reduction (CCR) is a commonly studied feature that consists of identifying the rate of simplification of final alpha-voiced stops in syllable-coda position, as in 'mis' for 'mist' or 'missed.' Although it occurs to different extents and under different constraints, it is a feature that occurs in all English dialects, making it a cornerstone variable for dialect comparison (Wolfram & Schilling-Estes 1998). Because it has been heavily studied[26] CCR remains a significant variable to trace when analyzing new dialect formation as it can indicate whether the variety in question is a continuation of a previous dialect studied, whether a dialect is accommodating to local dialects, or whether it is developing independently due to continued substrate effects (see Wolfram, Childs, & Torbert 2000). Further, because CCR functions under both phonological and syntactic constraints (Fasold 1972), comparing its patterns of usage across LOR and field sites with other clearly syntactic or clearly phonetic variables may reveal the nuanced differences in how phonetic and syntactic variables emerge in acclimatizing varieties. An added advantage to studying this feature is that as it is one of the few quantitatively analyzed morphosyntactic features studied in Latino English (Bayley & Santa Ana 2004), this data may be compared to previous studies to identify if varieties of Latino English in North Carolina are distinct from Latino English in other areas of the country. Of particular interest to this study is whether or not significant differences between field sites or LOR categories influence deletion patterns as such a correlation may indicate regional and acquisitional variation.

---

[26] E.g. Labov, Cohen, Robbins & Lewis (1968), Wolfram (1969), Fasold (1972), Wolfram (1974), Labov (1989), Santa Ana (1991), etc. . .

5.1 A Brief History of Consonant Cluster Reduction Analysis in Latino English

Due to the saliency of CCR, as well as the educational implications that result from past tense leveling created through the phonological processes that affect it, CCR is a trait that has been studied in-depth since the first quantitative variationist studies of Latino English (Wolfram 1974). Previous research of CCR demonstrates that Puerto Rican and some varieties of Chicano English share similarities with AAVE in that reduction is favored in all environments excepting prevocalic bimorphemic environments, especially when participants indicate increased ties to African American social networks (Wolfram 1974, Fought 2003). Increased deletion before a pause, an environment where consonants are typically retained in Anglo American English, is particularly salient (Wald 1984, Bayley 1994, Fought 2003). Such patterns may be reinforced in a community such as Durham, North Carolina, where the participants necessarily experience increased contact with native AAVE speakers (see section 2 for a discussion of community demographics). Notably, other studies indicate different patterns, including Santa Ana's (1991) analysis which found that Latino English in L.A. treated pre-pausal environments as vowels much like other Anglo Vernacular varieties[27]. One area where studies do agree is that Latino English generally demonstrates higher overall levels of reduction than most English varieties, although less so than Vietnamese English (Wolfram & Schilling-Estes 1998:253).

---

[27] It is worth noting that traditionally LA (124th highest percentage of African Americans in a metropolitan region according to the 2000 US census) has a much smaller African American population than New York City (64th highest percentage of African Americans in a metropolitan region according to the 2000 US census).

Other studies identify regional and generational differences as well as an increased impact of phonological constraints (Bayley 1994) and sonoric constraints (Santa Ana 1996). Consonant cluster reduction has also been analyzed in conjunction with social categories such as gender, where females were shown to reduce less so than their male peers (Hartford 1978). As the focus of this study is on emerging regional variation, the major social factors for analysis will continue to be length of residency and community

5.2 Methods

Tokens from all 36 participants were collected from the sociolinguistic interviews described in the second section of this paper (2. Methods). Although Spanish language tokens were excluded, many of the interviews used in this analysis did contain CS, as explored in chapter 3. Word final consonant clusters ending in alveolar stops which demonstrate uniformity in voicing were coded for analysis. Cluster of [sp] and [sk] were also analyzed following conventions in the field (Wolfram, et al. 2000). As shown in table 18 factors chosen for analysis followed convention as well, including the morphemic status of the cluster and the phonemic category of the following environment. The category of morphemic status refers to whether or not the cluster results from the addition of a stop to indicate a past-tense marker, a derived adjective, or a past participle (e.g. 'He pa*ssed* the class,' 'the relie*ved* student,' 'or 'The papers were pa*ssed* out') (Wolfram 1974). Deleting the final consonant in the cluster in such an environment can result in past-tense unmarking, which may explain why in most dialects this environment is more resilient in retaining the full cluster than other locations that do not encode such information. The immediate

following environment was coded as pause, vowel, or consonant. Filled pauses such as 'named *um*' were coded as vowels. Gender, LOR, and field site were also added as factor groups.

Table 18. CCR Factor Groups

Factor group	Factor Example	Factor Example	Factor Example
Consonant Cluster (Dependent Variable)	Reduced *'bes' frien''*	Full *'be**st** friend'*	
Morpheme Status	Mono-morphemic *'best friend'*	Bi-morphemic *'called* Erin'	
Following Environment	Consonant 'my girlfriend *says*'	Vowel 'my girlfriend *is*'	Pause 'Then we *moved.*'
LOR	0-3 yrs	4-8 yrs	$\geq$ 9 yrs
Gender	male	female	
Field Site	Hickory	Durham	

Semi-weak verbs, such as 'told' and 'found,' were excluded from the token count, along with any word occurring within a single participant's speech more than five times. The words 'and,' and 'just' were excluded from analysis due to their high frequency. If the presence or absence of the final consonant was unable to be verified aurally, a spectrogram analysis was performed using *PRAAT*. Following these procedures, 823 tokens of consonant clusters were extracted from the data for analysis: 466 coming from the Hickory data and 357 from the Durham data.

5.3 Results

As demonstrated in figure 17, overall reduction rates were high at 59 percent,
although this follows other findings which identify reduction rates for Chicano and Puerto
Rican English (Santa Ana 1991, Labov 1989, Wolfram & Schilling-Estes 1998). In-line with
previously-noted studies (see section 5.1), reduction in bimorphemic prevocalic
environments was low in comparison to other environments.
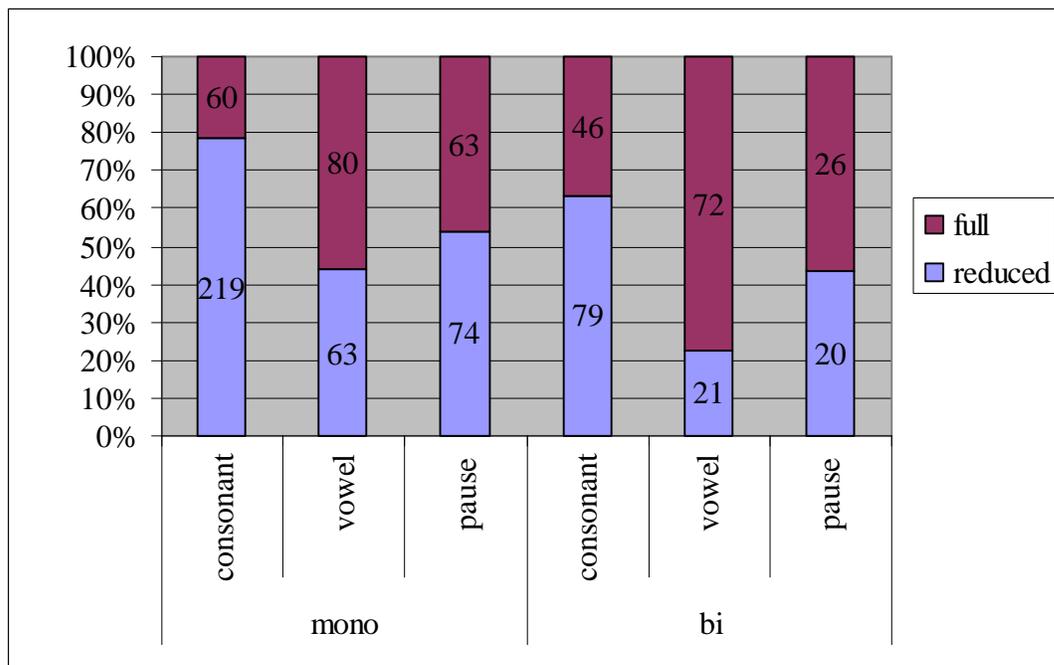
Figure 17. Total Token Reduction by Percent with Token Totals

To identify the interactions of field site and gender with CCR, a GoldVarb analysis
was constructed using the factors listed in table 18. The "best fit" model that emerged from a
step-up/step-down regression included site location, gender, morphemic status and following

environment (see table 19). Contrary to previous findings, the females in this study slightly

favored reduction. Again, this may be due to the nature of data collection as the females may

have felt more at ease, and therefore be more likely to use informal speech, with the female

interviewers who collected the data, although comparable male dyad data would be necessary

to test this hypothesis.

Table 19. CCR GoldVarb Run with Both Field Sites
Application Value = *Reduced* input = .59
Chi-square = 23.6272   chi-square/cell = 0.9845
Log likelihood = -489.968

Factor Group	Factors	Factor Weight (Input and Weight)
Site Location	Hickory	.412 (.50)
	Durham	.614 (.70)
Gender	Female	.543 (.63)
	Male	.446 (.54)
Morphemic Status	Monomorphemic	.556 (.64)
	Bimorphemic	.383 (.47)
Following Environment	Consonant	.68 (.75)
	Vowel	.279 (.36)
	Pause	.392 (.48)

This analysis also indicates that differences in reduction levels between field sites are

significant (see figure 18). Durham speakers appear to consistently reduce consonant

clusters at higher rates than Hickory speakers, except for in monomorphemic pre-consonantal

environment. Particularly striking is the difference in how participants from each field site

treat pre-pausal environments. Durham participants significantly reduce consonant clusters

more often than Hickory participants in pre-pausal environments regardless of morpheme

status.

| | | mono | | | bi | |
	c	v	p	c	v	p
☐ Hickory	0.8	0.32	0.36	0.58	0.2	0.23
☐ Durham	0.77	0.57	0.67	0.72	0.28	0.63
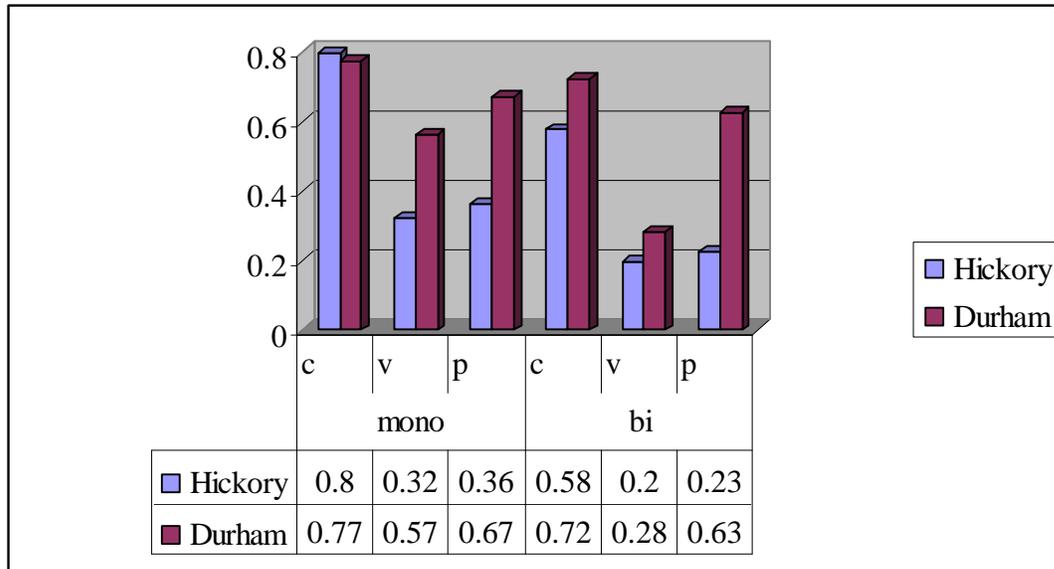
Figure 18. Percent Reduction by Field Site and Morphosyntactic Constraints
Difference in monomorphemic pre-pausal environments is significant at $p < .001$, chi-square $= 12.8$.
Difference in bimorphemic pre-pausal environments is significant at $p = .007$, chi-square $= 7.39$

Although LOR was included as a factor group for analysis when designing the

statistical model, this variable was thrown out in the step-up/step-down regression which

established the "best-fit" model. However, aggregating the data may obscure more subtle

differences between the participants with shorter LORs and longer LORS. To explore the

impact of LOR on reduction, tokens were separated out by LOR group in figure 19.
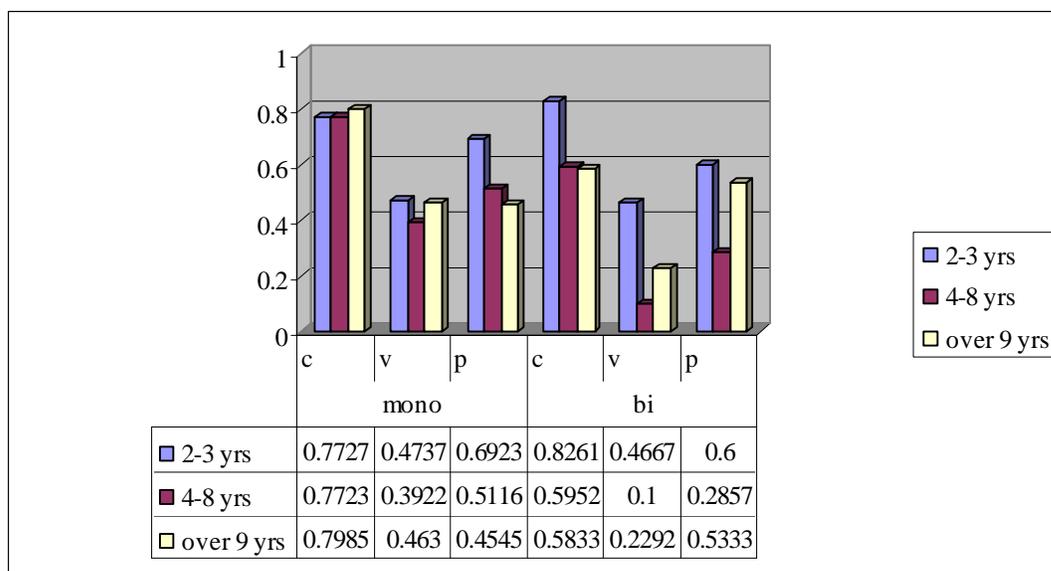
Figure 19. Percent Reduction by LOR and Morphosyntactic Constraints

Table 20. Chi-Square Runs for CCR Differences across LOR Groups

Morpheme Status		P Value	Chi-Square	Significant
Mono	Consonant	P = .869	.281	Not significant
Mono	Vowel	P = .683	.764	Not significant
Mono	Pause	P = .067	5.40	Not significant
Bi	Consonant	P =.101	4.58	Not significant
Bi	**Vowel**	**P =.021**	**7.70**	**Significant**
Bi	Pause	P =.165	3.60	Not Significant

Out of all the environments explored by the analysis, the only significant difference across LOR groups is in the bimorphemic prevocalic environment. This may indicate an early establishment of community norms – by roughly the third year. Because reduction of consonant clusters before a pause is variable across dialects, but prevocalic reduction is largely disfavored cross-dialectically (Wolfram, et al. 2000, Labov 1989), the tendency to preserve these clusters among those with more exposure to native-speaker varieties of English may result from the consistency of this constraint in American English. However, high levels of reduction remain in all other environments, possibly as a result of substrate

influences (Santa Ana 1996). Impressionistically, the second LOR group once again appears to favor the prestige variety by retaining clusters in bimorphemic positions more frequently before pauses and vowels, although this difference is only significant in the bimorphemic prevocalic environment.

Because the treatment of prepausal environments is significantly different across field sites, and because the constraining effect of this environment frequently varies across social and ethnic groups, with Anglo varieties typically retaining clusters before a pause while African American varieties tend to reduce in this environment (Wolfram, et al. 2000), a closer look at these tokens is warranted. In figure 20, percent reduction across LOR groups again shows a u-shaped pattern with the middle group demonstrating the least amount of reduction. Although this pattern is consistent across field sites, the Durham groups demonstrate consistently elevated reduction levels, following the patterns found in African-American influenced varieties. As with the analysis of quotative frames, this demonstrates that the reduction pattern that appears in the data cannot be solely attributed to acquisitional patterns. Even at this early stage in dialect formation, local norms emerge.
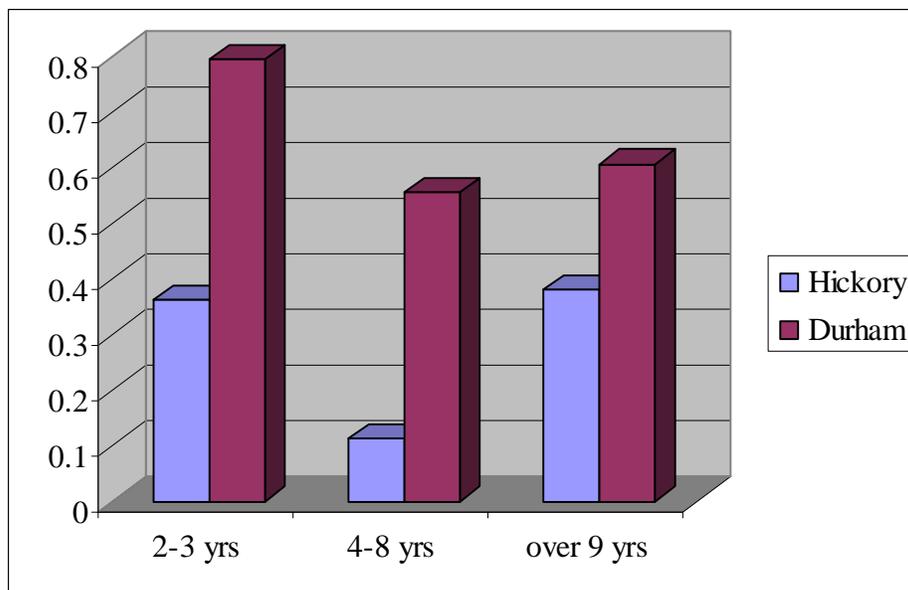
Figure 20. Percent Reduction of Prepausal Tokens across LOR and Field Site

5.4 Discussion

While past tense construction for *say* and *be like* were clearly influenced by language learning, as demonstrated by the statistically-significant differences in tense usage across LOR group, the fine phonological processes involved in consonant cluster reduction may prove more lasting within the dialects under examination. Santa Ana (1996) speculates that maintained high levels of deletion are linked to the more syllable-timed prosody of Latino English, a Spanish substrate variable. Syllable-timing[28] is said to favor CV constructions and disfavor consonant clusters. Because Spanish is a more syllable-timed language it is said to disfavors clusters. As prosodic substrates are known to endure within dialects (Thomas, Carter & Cogshill 2006), this effect may continue to influence consonant cluster reduction rates in all but the most syntactically and aurally salient environments. However, what

---

[28] Syllable refers to languages where syllables receive roughly equal stress as opposed to the stress being dispersed across feet as occurs in stress-timed languages in English.

emerges as salient depends on the community in which the speaker is exposed to the variable. So, prepausal environments may emerge as salient in the largely Anglo community of Hickory, while not so in Durham where a large AAVE speaking population likely favors deletion in this environment. Thus, processes of *acclimatization* are locally-specific, as demonstrated by the striking patterns in this data.

Even though Latinos represent the largest minority group in the United States (US Census 2006), there are few acoustic analyses of Latino English vowel systems (Godinez 1984, Godinez & Maddieson 1985, Veatch 1991, Thomas 1993 & 2001, Fought 1999 & 2003, Roeder 2006). Further, those studies that do exist[29] usually describe long-standing Mexican American communities in the Western and Southwestern regions of the United States. In order to more thoroughly understand regional ethnic variation, comparable studies must be undertaken in other areas of the United States.

And, while research has begun on the varieties of Latino English that are currently being established in the Mid-Atlantic region of the United States (Moriello 2003; Wolfram, Carter & Moriello 2004; Thomas, Carter & Cogshall 2006; Carter 2007), acoustic vowel analysis has been largely restricted to specific salient variables such as the /ai/ diphthong, prenasal /æ/, /o/, and /u/ (Carter 2007, Wolfram, et al. 2004, Moriello 2003).[30] Although such studies can demonstrate whether or not Latino populations in the region are accommodating to specific local norms, a more extensive analysis is necessary for comparison to Latino English research in other regions of the United States, such as California and Texas. Further, as Thomas (1993) has shown, such a study may identify other variables that deserve attention.

This study attempts to enhance current knowledge of Latino English by providing a spectral analysis of the vowel spaces of four Latino males, two from each field site under

---

[29] Non-acoustic studies both within and outside of the Southwest are more common (e.g. Poplack 1978, Hartford 1978, Gordon 2000).

[30] There is one full vowel chart of a North Carolina Latino resident in Thomas (2001).

investigation. A selected analysis of /æ/ and /o/ in the speech of six additional participants, for a total of ten speakers, is compared with local community members. This full and selected analysis may demonstrate how vowel systems emerge within Latino English in Hickory and Durham, North Carolina, further indicating how the varieties acclimate within these communities.

6.1 Background

Past acoustic vowel analyses for Latino English speakers demonstrate that while Latino English in the Southwest and California show Spanish substrate influences, it both differs from the language of speakers acquiring English (e.g. Santa Ana 1991) and from local Anglo varieties (e.g. Veatch 1991, Godinez 1984, Godinez & Maddieson 1985, Santa Ana 1991). Specifically, Godinez & Maddieson (1985) found that front vowels of Californian Latino English, including /I/, /ɛ/, and /æ/, were generally higher than in Anglo varieties, while back vowels like /Y/ were generally more backed. Weakened glides on the English diphthongs /e/ and /o/ (Fought 2003, Thomas 2001, Santa Ana 1991) and mergers (Thomas 2001, Bernstein 1993) are also frequently noted by researchers. Compared to African American English, a variety that is known for retaining back vowels (e.g. Thomas 1989), /o/ in Latino English appear even more backed than in AAVE (Thomas, Carter & Cogshall 2006).   However, there have been contradictory findings demonstrating increased accommodation to local norms, including /u/ fronting in California (Fought 1999, 2003). That Latino English speakers tend to resist /æ/ lowering and backing is similarly contentious in the literature, with some maintaining that Latino English resists backing (e.g. Thomas

82

2001, Godinez 1984), while others find evidence of accommodation to local variants (Fought 2003).   More studies demonstrating accommodation to regional norms include Roeder (2006), who identifies accommodation to certain features of the Northern City Vowel Shift among females under the age of 25 in Michigan, and Thompson (1975), who found /ai/ glide weakening among some Latinos in Austin, TX. In summary, previous studies find both difference and accommodation when comparing the variety to local norms.

Acoustical studies of Latino English outside of Texas, Michigan, and California are less common. More recent studies of Latino English in North Carolina exhibit that the variety may have much in common with Latino English from established communities in Texas, including backing of /o/, increased syllable timing compared to Anglo varieties, and resistance to prenasal /æ/ raising (Thomas, et al. 2006). While Wolfram, et al. (2004) found evidence of some lexicalized accommodation, such as weakened /ai/ glides for "five" and "North Carolina," the majority of the data patterns intermediate to Spanish and local Southern varieties. The one study that did find local accommodation in a young girl from Cary, North Carolina (Carter 2007) found that such accommodation dissipated when the participant graduated from a majority Anglo elementary school to a middle school that had a larger Hispanic population.

While evidence of accommodation to local norms for Latino speakers in North Carolina has been limited, a full vowel analysis may reveal areas of accommodation overlooked in previous studies. Further, the studies listed above analyzed the speech of Latinos living in Raleigh, North Carolina and Siler City, North Carolina. As discussed in the introduction to this analysis, the influx of individuals from diverse regions in Raleigh would prohibit acquisition of Southern English norms, while the density of the Latino population in

83

Siler City may make that community particularly resistant to accommodation due to insularity[31]. As Hickory is mid-sized town with a largely Anglo population and Durham is a mid-sized city with a large African American population, and, as both have Latino populations that are similar to North Carolina as a whole, these field sites may offer a more general picture of how Latino speakers are accommodating to local speech varieties in other areas of the Mid-Atlantic South.

6.2 Methods

Vowel tokens of /ij/, /I/, /e/, /ɛ/, /æ/, /æn/, /Λ/, /A/, /O/, /ow/, /ol/, ^{K}u/, /^{T}u/,/ ul/, /Y/, /Yl/, /ai/, /au/, /oi/, /ir/, /ejr/, /ar/, /or/, and /r/ were taken from a selection of the previously-discussed sociolinguistic interviews (see table 21). All speakers chosen for this section have an LOR of at least 5 years. Additional tokens of /æ/, /æn/, and /o/ were taken from three participants in each field site so that these vowels could be compared across ten individuals with the community comparison participants (bolded in the chart), as listed in table 21.

---

[31] According to the 2000 United States Senate, 39 percent of Siler City reported Latino ethnicity. Meanwhile, Wake County, where Raleigh, North Carolina is located, grew by over 10 percent between 2000 and 2003, thus indicating the relative mobility of its population. An estimated 54 percent of the population currently living in Raleigh, North Carolina, were born outside of North Carolina.

Table 21. Select Participants for Vowel Analysis

Participant	Field site	Type of Analysis	Age	Ethnicity
Jesus	Durham	Full	13	Latino
Josue Y.	Durham	Full	11	Latino
Miguel	Hickory	Full	16	Latino
David	Hickory	Full	14	Latino
Elma	Hickory	Partial	9	Latino
Jasmine	Hickory	Partial	9	Latino
Steven	Hickory	Partial	9	Latino
Jalinda	Durham	Partial	12	Latino
Marcela	Durham	Partial	8	Latino
Jackson	Durham	Partial	10	Latino
Yolanda[32]	Hickory	Partial	Adult	Latino
Jenny	**Hickory**	**Partial**	**20**	**Anglo**
Nick	**Hickory**	**Partial**	**17**	**African American**
Amy	**Durham**	**Partial**	**12**	**African American**

570 vowels were measured for the full vowel charts. Diphthongs were measured .35 milliseconds from each end of the vowel space, while monophthongs were measured at the center. The fundamental frequency, duration of the vowel, and first through fourth formant were taken for each token using PRAAT. Due to co-articulatory effects tokens that appeared before or after a nasal, lateral, or /r/ are excluded from analysis or grouped separately for comparison and marked with a superscript within the vowel charts following Thomas (2001). The mean of the first two formants was taken for all appropriate measures of each vowel. Vowels were then plotted with the second formant on the x axis and the first formant on the y axis.

All tokens for the partial vowel analysis were measured from the center of the vowel, and similar environments were avoided to prevent skewed measures resulting from co-

---

[32] This participant represents a speaker who learned English after the critical period.

articulatory effects. Following these guidelines, a total of 230 additional tokens were measured, with 96 coming from community representatives as listed in table 21 and the remaining 134 coming from the other Latino participants. Generally, at least six tokens of each vowel were measured from either the sociolinguistic interview or the reading exercise for each individual, although some speakers had fewer tokens for /æn/ and /o/. Again, the mean of the first three formants was taken for all appropriate measures of each vowel. These means were then normalized through the BARK[33] Difference Metric (Syrdal & Gopal 1986) which arrives at the values Z3-Z1 (BARK-converted F3-BARK-Converted F1) to indicate height and Z3-Z2 (BARK-converted F3-BARK-Converted F2) to indicate horizontal placement.

6. 3 Results

6.3.1 Full Vowel Space Analysis

Josue was eleven and had lived in Durham, North Carolina for six years at the time of the interview. He reported having friends from Taiwan who spoke English as their second language, as well as other Latino Spanish/English bilingual friends. In his spare time, Josue stated that he enjoyed video games, Japanese cartoons, and listening to the local Mexican radio station. Vowel tokens for Josue came from two interviews, one which was 36.5 minutes long (dps029 in the NCSLAAP data base) and another that was 42 minutes long, approximately half of which was conducted in Spanish.

---

33 The following formula was used to arrive at Z values: $Z_i = 26.81/(1+1960/F_i) - 0.53$ (Traunmüller 1997).

Josue's speech (Figure 21) largely mirrors Latino English found in Texas as recorded by Thomas (2001). Prenasal /æ/ does not appear to be raised. Also, while mergers are prevalent in his system (/o/= /O/, /or/=/Or/), he does not monophthongize /i/or /e/. His front vowels do not appeared to be dramatically raised either. His back vowels, including /Λ/[34], however, are not fronted and his /o/ and /u/ tokens demonstrate weak glides. Unlike Siler City Latinos (Wolfram, et al. 2004), Josue does not have raised nuclei of /ai/ diphthongs, although his glides are strong. His /au/ diphthong is low and resists fronting, a trait that is common to both AAVE and ChE (Thomas 2001:26). In short, Josue's speech has a few features which are not characteristic of other descriptions of Latino English speech, such as resistance to front vowel raising, but is generally similar to other Latino speakers in his high number of mergers and resistance to fronting backed vowels (Thomas 2001:150, Godinez & Maddieson 1985).

---

[34] This is an interesting trait as Thomas (2001:18) noted that backed wedge is not necessarily common in Chicano English, but may be more common in among those who speak English as a second language.
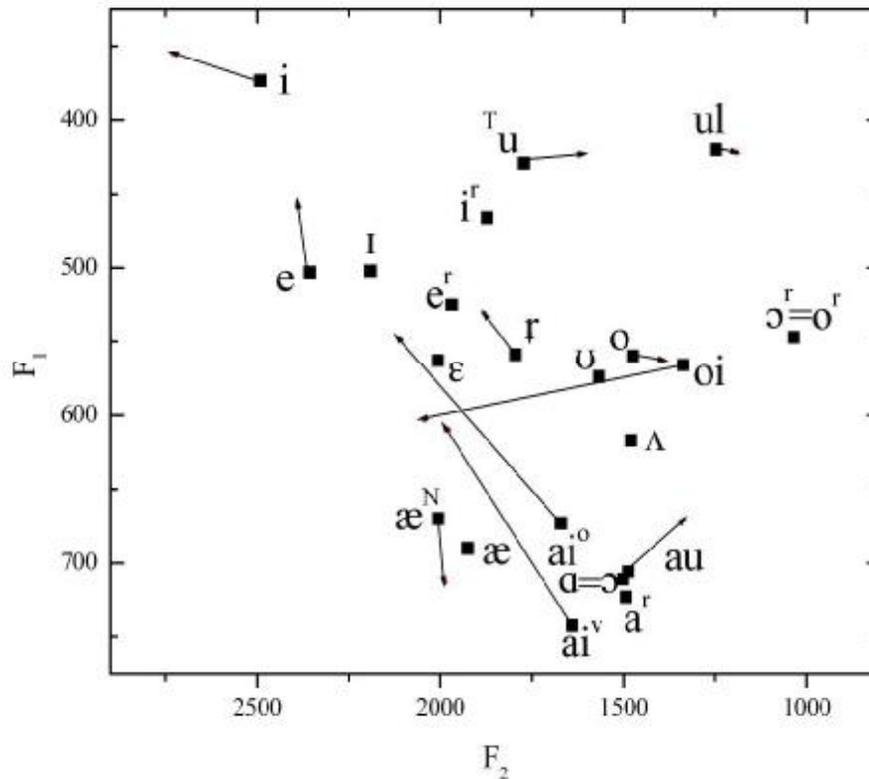
Figure 21. Formant Plot of Latino Male, 11 yrs Old, in Durham, NC

Jesus was thirteen at the time of the interview and had lived in Durham, North Carolina for eight years. He self-reports as a gang member with three of his four brothers also participating in gang activity. Jesus has family in rural Georgia who he occasionally visits, including his oldest brother and cousins. Vowels were measured from a 51 minute interview conducted at school in the presence of peers.

Jesus (Figure 22.), like Josue and following expected norms, also has several mergers in his vowel space, including /o/= /O/ and /or/=/Or/. Although /ul/ is distinct from /Yl/ this may be due to the small number of tokens available in the data. Further, Jesus has a fronted raised /au/ diphthong, which is said to be common in many regions of the United States,

especially the South (Thomas 2001:27). He differentiates between pre-voiced and pre-voiceless /ai/, although not in the expected manner for accommodation to Southern American English varieties.[35] There appears to be some lexicalized glide weakening of the diphthong in words like "five," a trend noted in Wolfram, et al. (2004).
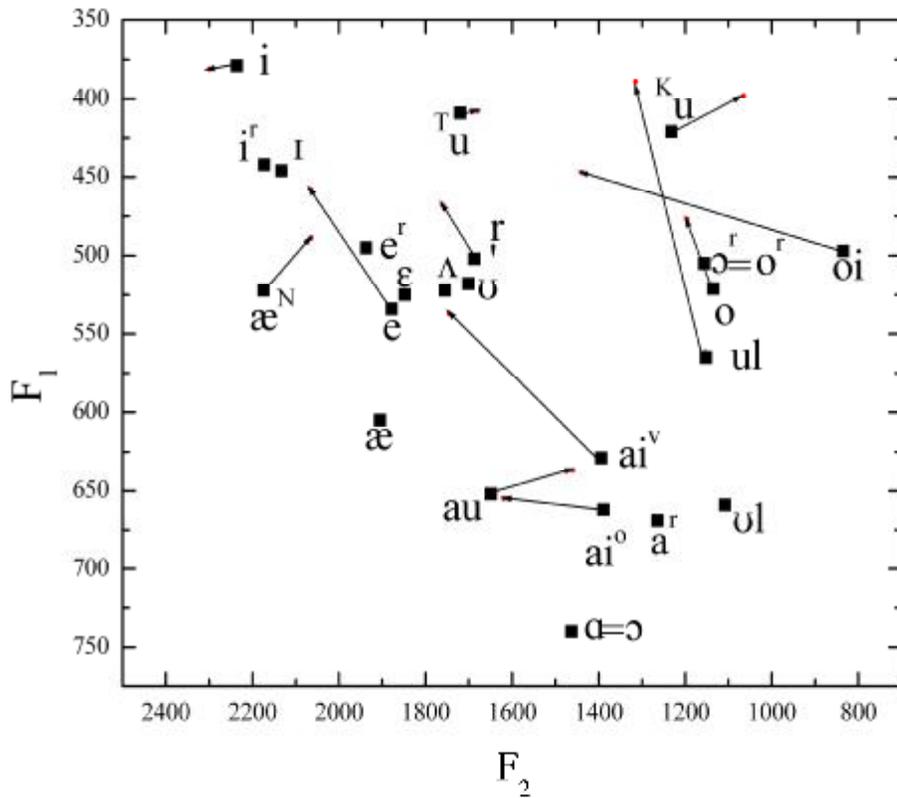

Figure 22. Formant Plot of Latino Male, 13 yrs Old, in Durham, NC

Although Jesus differentiates prenasal /æ/ from non-pre-nasal /æ/, all of his front vowels are raised, a feature that Godinez & Maddieson (1985) list as distinctively Latino. And, while /e/ is not monophthongal, /i/ shows only a weak glide. This is also reflected in

---

[35] It should be noted that the /ai/ measurements were based on four tokens for both pre-voiced and pre-voiceless environments, which may have skewed the data.

his raised /ai/ nuclei. He variably weakens /r/, something that is especially common in his pronunciation of /ar/. This is an unexpected feature within Latino English and may represent local accommodation as *r*-lessness is common in various Southern varieties, as well as African American English (Thomas 2001:31), although a comparative acoustic analysis would be necessary to verify the source of accommodation.
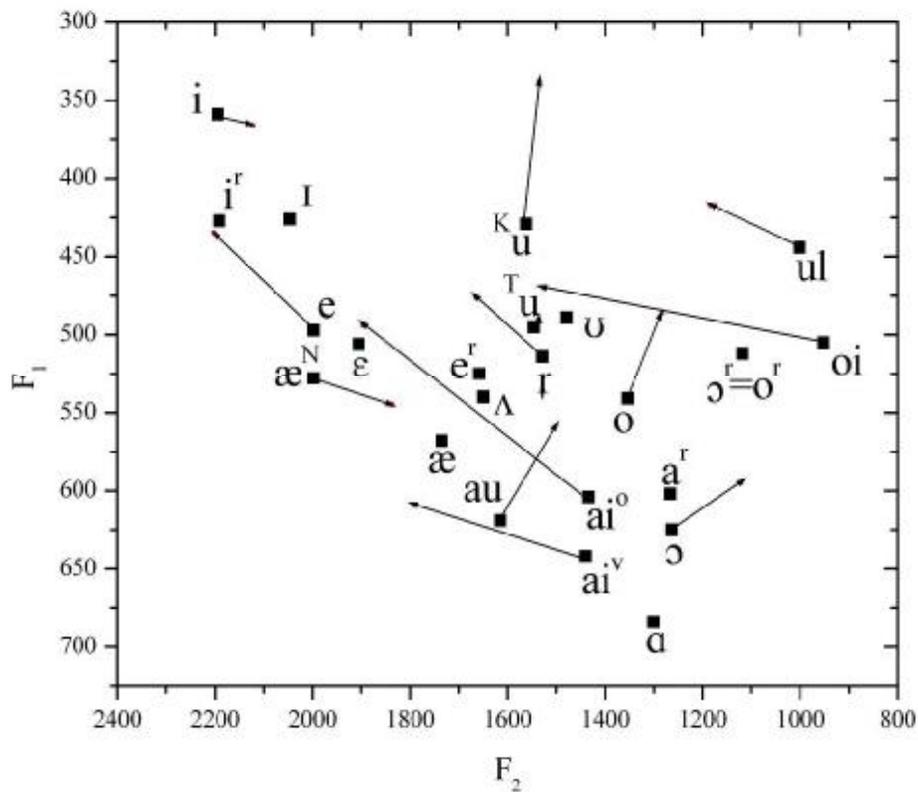

Figure 23. Formant Plot of Latino Male, 14 yrs Old, in Hickory, NC

David, who was fourteen at the time of the interview, also lived in Hickory, North Carolina. However, he was born in Los Angeles, California. David attends the local Catholic Church, and receives tutoring from La Comunidad after school. David listens to R

& B music and plays racing video games in his free time. He wants to be a federal agent or a sheriff when he grows up. Vowels were measured from his fifty-four minute interview recorded in a classroom at La Comunidad.

David, as shown in figure 23, also differentiates prenasal /æ/ from non-pre-nasal /æ/ while raising his front vowels. He also, like Jesus, has a fronted /au/ diphthong. The glide on /^{K}u/ may be a lexicalized tendency for the token "food." There appears to be a slight raising of the nuclei of /ai/ in both environments; but, more surprisingly, David appropriately distinguishes between voiceless and voiced environments. Although this distinction is represented by limited tokens of "inside" and "Friday," they are not tokens that are noted for a tendency to be lexicalized in Wolfram, et al. (2004). Further, David does not merge cot/caught. As both /ai/ ungliding and maintenance of the cot/caught distinction have been listed as Southern and African American features – but not as features of Latino English[36] - these two findings are surprising. Another distinction between Jesus and David is that David has more fronted realizations of his back vowels. There is a weak glide on his /i/ and /^{T}u/ token, but /e/ and /o/ are not monophthongal. While David is *r*-ful, his weakened /ai/ glides and cot/caught distinction make him sound rather more "Southern" than the first two participants analyzed.

Miguel from Hickory, North Carolina was sixteen at the time of the interview in 2006. Vowel tokens were taken from his sixty-one minute interview recorded in a classroom at La Comunidad. His LOR in the United States was eight years at that time, although he occasionally traveled back to Mexico, where he was born, to visit his grandparents. Miguel reported that he enjoys watching shows such as "House," a medical comedy, and "Bones," a

[36] Although it should be noted that the cot/caught distinction was present in the speech of four Mexican Americans in Thomas's (2001) study as well as noted by Veatch (1991) in Los Angeles.

forensics drama, during his free time. He wishes to study architecture in Mexico in the future

and currently volunteers at La Comunidad after school.
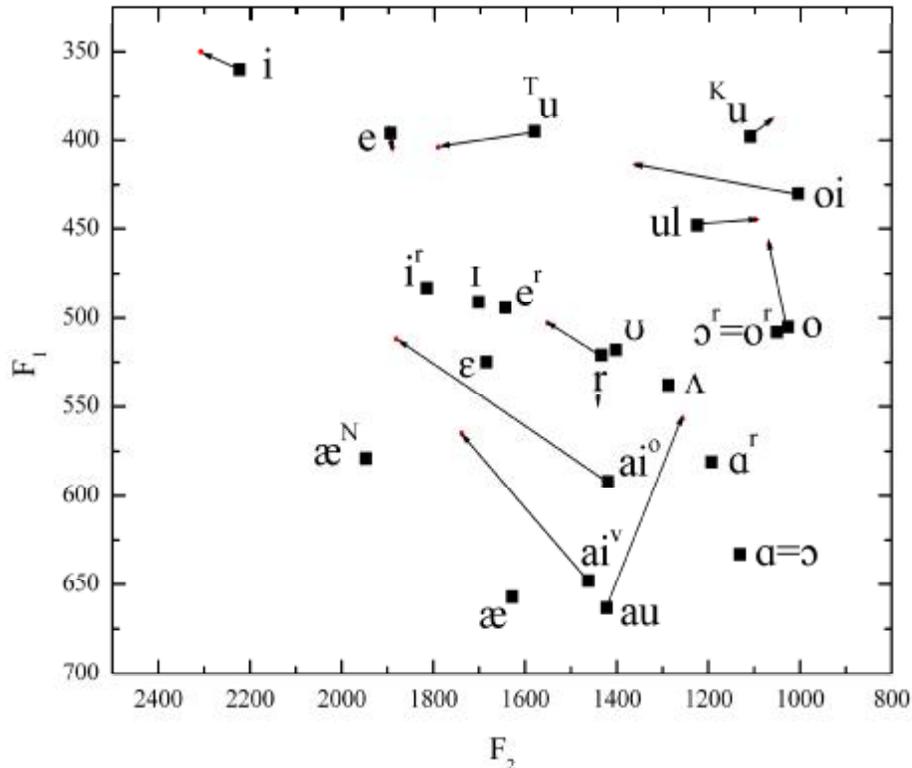
Figure 24. Formant Plot of Latino Male, 16 yrs Old, in Hickory, NC

Miguel (figure 24) reverses some of David's patterns in that his /e/ and /i/ have weak

glides while his /o/ and /u/ are more glided. Following David and Jesus, Miguel also

differentiates prenasal /æ/ from non-prenasal /æ/. His non-pre-nasal /æ/ is considerably

lowered which found to be a trait of young Anglo speech in various parts of the US,

especially Ohio and the West Coast (e.g. Thomas 2001, Luthin 1987), although once again a

comparative acoustic analysis would be necessary to verify the source of this feature. He

does have the cot/caught merger, with this vowel occurring towards the low back section of

his vowel space. This is noted by both Thomas (2001) and Fought (2003) as an Anglo feature. Unlike David and Jesus, Miguel's /I/ appears more central. That being said, /e/ is quite raised. Miguel's back vowels also show similar variation when compared with previously noted trends. While /Y/ appears fronted, /o/ remains backed. While not sounding "Southern," Miguel does tend to sound leveled.

The pattern that arises from the comparison of these charts is that each speaker demonstrates at least some previously noted Latino English features, as summarized in table 22. However, surprising differences from the previous literature occur, especially in the vowel spaces of the individuals from Hickory, North Carolina. For example, three of the four speakers differentiate prenasal /æ/ from /æ/ occurring in other environments, which differs from other regional findings (Thomas, et al. 2006). While Jesus and David raise their front vowels, Miguel and Josue only raise /e/. Mergers are prevalent, but not categorical, as demonstrated by David. And, although Wolfram, et al., stated that /ai/ glide weakening was largely lexicalized, at least David has conditioned glide weakening before voiced consonants.

Table 22. Selected comparison of vowel features

Participant	Pre-nasal /æ/	Non-pre-nasal /æ/	Vowels with weakened glides	Cot-caught Merger	Back Vowels
Josue	Not Raised	Not Raised	/o/ and /u/ But not /i/ or /e/	Yes	All but /Tu/ are backed
Jesus	Raised	Raised	/i/ but not /e/ or /o/	Yes	/Y/, /Λ/ are fronted, others are backed
David	Raised	Raised	/i/ but not /e/ and /o/	No	/o/, /Y/, and /u/ are fronted
Marcos	Raised	Backed	/e/ and /i/ but not /o/ and /u/	Yes	/Y/ is fronted, /o/ is backed

Both community makeup and individual orientation may influence such accommodation. Miguel's low back merger may be related to his orientation toward mainstream Anglo culture, as demonstrated by his tendency to watch T.V. shows marketed towards such audiences. As Josue identifies with a multi-ethnic peer group who largely speaks English as a second language, this may encourage mergers and leveling. David, on the other hand, lives in Longview, a neighborhood with over 57 percent African American population (Figure 25).

Longview – a neighborhood of Hickory, NC

Data Classes

Percent
- 0.7 - 3.4
- 4.2 - 6.1
- 8.3 - 13.5
- 24.0 - 24.0
- 57.6 - 57.6

Features
- Major Road
- Street
- Stream/Waterbody
- Stream/Waterbody

Approx. 20 miles across.
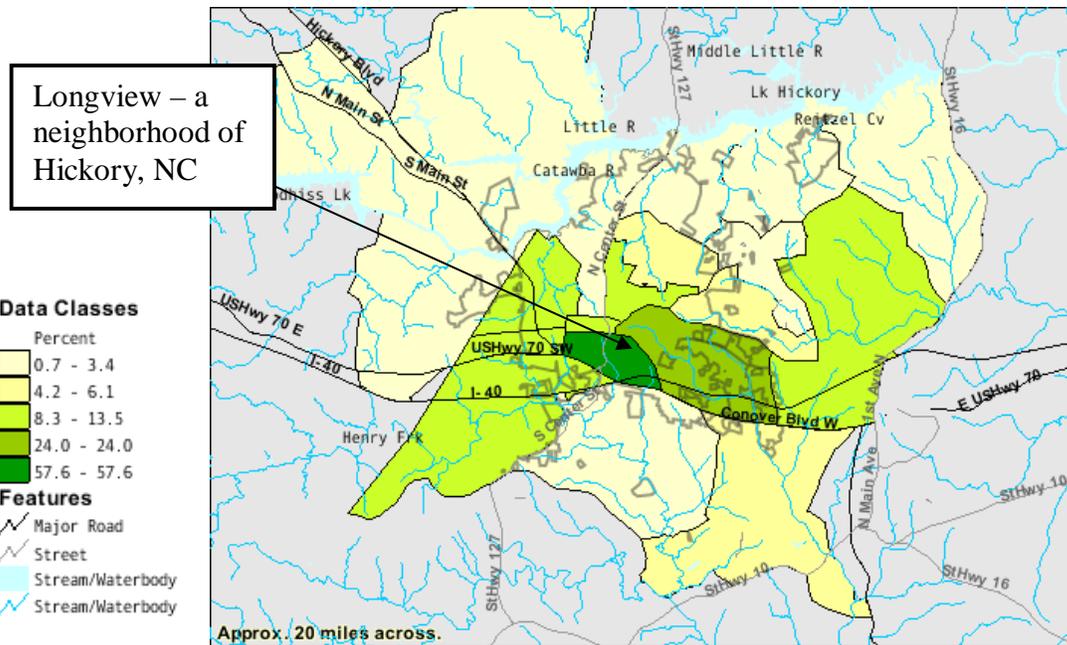
Figure 25. African American Population Density of Hickory, North Carolina

As such, his conditioned /ai/ weakening and retention of the cot/caught merger may be due to community influence. Similarly, as Jesus lives in Durham, North Carolina, with its large African American population, community demographics may at least partially explain his variant /ai/ glide weakening and *r*-lessness.

6.3.2 /æ/ and /o/

As /æ/ and /o/ show variation among these four speakers, and as there is a comparable study completed with Raleigh, NC and Pearsall, TX Latinos, these two variables were singled out for further investigation. The results of the plotted normalized vowels are given in figure 25.
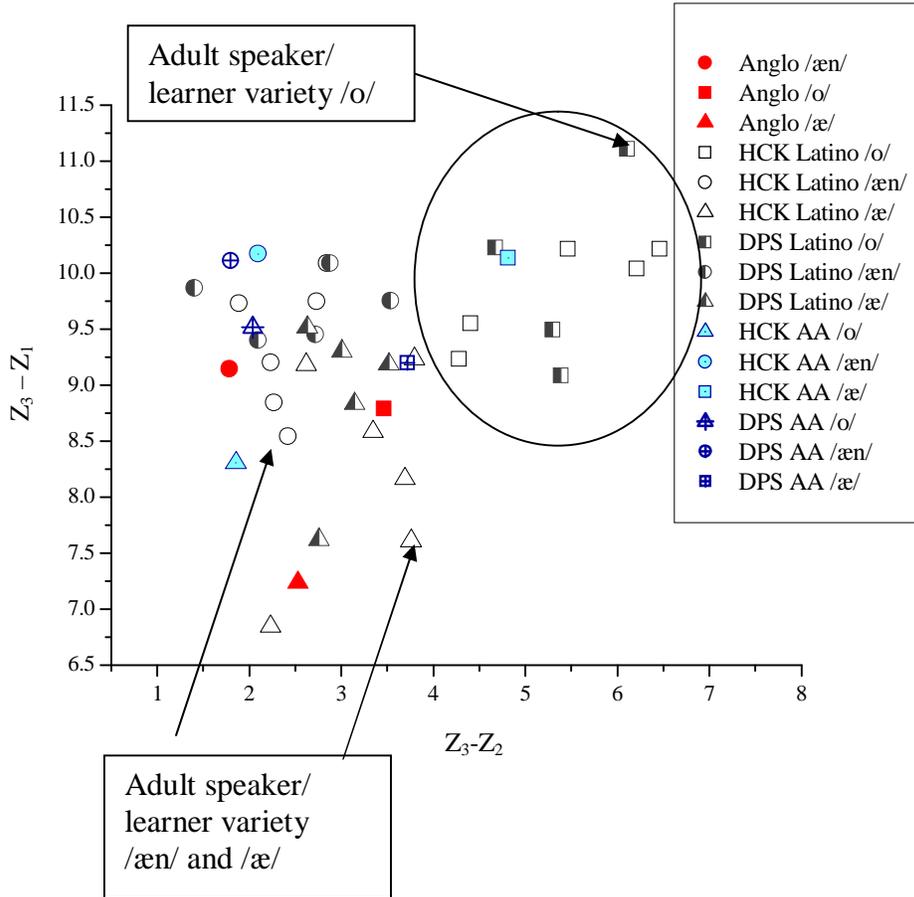
Figure 26.   Comparison of Normalized /æ/ and /o/ Values

Initial observations show that all of the Latino participants have more backed /o/
tokens than either the Anglo community member from Hickory or the African American
community member from Durham, which aligns with previous findings of Latino English in
North Carolina (Thomas, et al. 2006). Both /o/ and non-pre-nasal /æ/ tokens have wide
pockets of variation, while pre-nasal /æ/ appears more concentrated.   Initially, Durham /o/
tokens appear more fronted, but this difference is not found to be statistically significant
when the groups were compared using a t-test (p = 0.61). Non-pre-nasal /æ/ looks to be more
raised in Durham than in Hickory, but this difference also comes out as insignificant using a

96

t-test (p = 0.28). However, this may be due to Marcela's tokens, whose non-pre-nasal /æ/ is much lower than her Durham compatriots. The tendency for /æ/ to be raised in Durham may be a combined effect of Spanish substrate influence and reinforcement from AAVE dialect norms, where /æ/ usually appears higher, closer to /ɛ/, than in Anglo varieties (Thomas 2001).
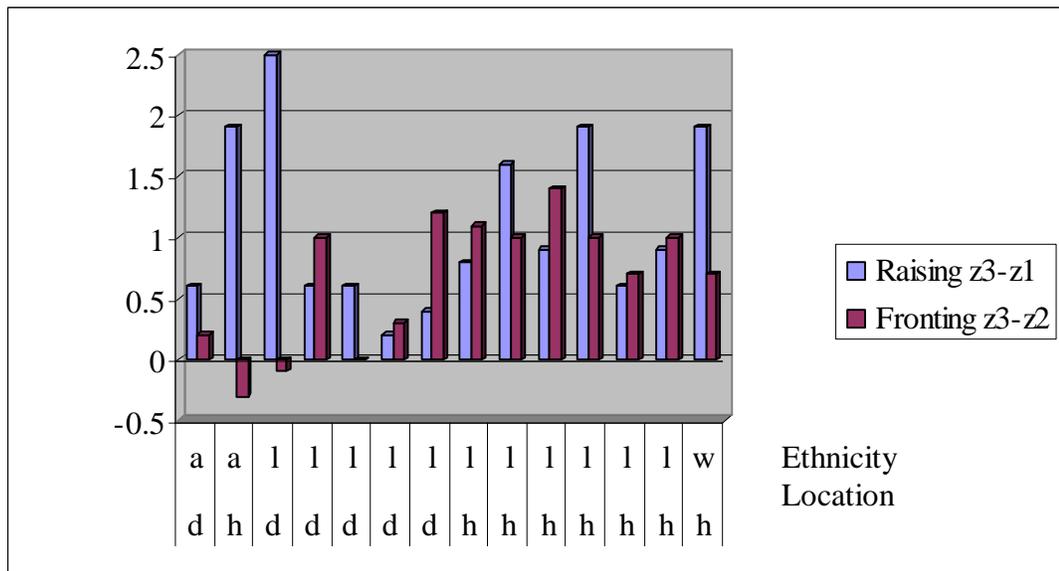

Figure 27. Pre-nasal /æ/ Raising by Ethnicity and Location

In order to analyze pre-nasal /æ/ raising, the normalized values for this environment were subtracted from non-pre-nasal /æ/ and plotted in figure 27. Across the board all participants differentiate to some degree between the two environments, although to different extents. The Durham Latinos appear to have wider pockets of variation among the group, which may be attributed to variable non-pre-nasal /æ/ raising, as raising in this environment would constrain the amount of raising that could occur in pre-nasal environments. Speakers like Jalinda provide a case-in-point. While her pre-nasal /æ/ tokens are only .6 Z higher than

her non-pre-nasal /æ/, both hover within 9 Z on the BARK Difference Metric. Further raising of these tokens would interfere with /I/ and /i/. This differs from speakers like Marcela and Emma who have considerably lower non-pre-nasal /æ/ tokens, around 7 Z in the BARK Difference Metric, thus allowing for increased raising of pre-nasal tokens, which is exactly what occurs.

6.5 Discussion

In some ways this research reinforces previous findings in that many features cited as common in Latino English in Texas and California also appear in the speech of these North Carolinian informants (Wolfram, et al. 2004, Moriello 2003, Thomas, et al. 2006, Carter 2007). Yet, the most striking pattern within the data is the amount of individual variation, as demonstrated by the lack of trends across the full vowel charts, as well as the wide vowel pockets in the selected analysis. Although there appear to be subtle community differences, such as /æ/ raising in Durham (which could be explained by accommodation to AAVE), these tendencies are not statistically significant.

Several explanations could account for such diversity. One possible explanation is that such variation is common, but goes unnoticed due to tendencies toward the exploration of aggregate data in the field (e.g. Johnstone 2000). Although I performed a thorough analysis of several speakers, the small number of participants included in this analysis enhances the appearance of differences that would appear less significant within larger aggregate data sets. Further, the participants chosen for this study represent an array of social identities, from gang members to honor role students. These identities necessarily lead

to different choices in social contacts, as well as in linguistic variables (e.g. Bucholtz 1999). Increased consistency may have been found by selecting individuals with similar peer groups, but this would have falsely created the appearance of a homogenous speech community.

Another explanation is that the emerging nature of the community naturally contributes to the heightened amount of variation found in the speech of the participants. The communities under analysis are exposed to a wide variety of dialects, including learner varieties, local Anglo varieties, local ethnic minority varieties, and Latino varieties from other regions of the country brought in by intra-national migration. The increased number of variants available to the individuals in these communities creates an environment ripe for koineization (e.g. Trudgill 1986, Thomas 1997, Trudgill, et al. 2000, Mesthrie 1993, Kerswill & Trudgill 2005, etc. . .). Initial stages in these processes involve heightened levels of variation, which eventually level out to create community norms among future generations. While such variation does not appear so extreme within the consonant cluster reduction analysis, vowel space does not benefit from the regulatory influence of prosodic structure in the same way syllable structure does. The nature of vocalic variables in English, which are rich in dialectal variation, encourages high levels of variation within emerging communities.

These explanations are not mutually exclusive, as the rich source of variation found in an emerging dialect community could indicate a larger pool of resources for the construction of individual linguistic identity. Further, the literature from California, Michigan, and the Southwest supports the assertion that vocalic systems in Latino English are variable and often correlate with class, gender, region, education levels, and social affiliations (e.g. Roeder 2006, Fought 1999, Hartford 1978, Thompson 1975). When explored next to

morphosyntactic variants, vowels may be more flexible to local and/or social influence. These variables will be interesting to track over generations to see if variation dies down or continues.

Exploring a wide range of variables within two distinct communities in the same region allows for a more complete understanding of how a variety acclimates to a new environment. Rather than broad generalizations regarding regional and national norms, such an analysis allows the observer to identify community patterns in the context of regional trends, informing how such local variation relates with previous findings from other areas of the country. The diversity of the variables included within this study also allows for an increased understanding of the different ways in which syntactic and phonetic variables acclimate to a new environment in the context of global trends. Using this technique of spatially situating syntactic and phonetic variation in an emerging community advances the general understanding of how new communities establish ordered variation as the features under analysis were shown to follow macro-trends while exhibiting locally specific norms.

Part of this approach includes analyzing the speech of individuals with a broad range of lengths of residency as to appropriately reflect the linguistic diversity found within an emerging community. Across variables where length of residency was considered a primary factor for analysis, significant differences were found between LOR groups. This is not surprising as increased exposure to English would favor increased proficiency. As expected, LOR influenced ordering of constraints within CS patterns, quotative frame patterns, and patterns in consonant cluster reduction. Patterns associated with sensitivity to tense-marking, such as retention of bimorphemic consonant clusters before a vowel, or marking of past-tense on quotative say, appear within the second and third LOR group. Similarly, sensitivity to monolingual norms may constrain the patterns of CS associated with the second and third

LOR group as demonstrated by the favoring of lexical insertion over other types of switches. These patterns reveal correlations in the acclimation process across linguistic variables and field sites.

For the purposes of analyzing emerging dialects, the variation within LOR groups as compared across field sites may demonstrate how locally-specific variation emerges early within communities. This is particularly true with the quotative frame and CS analysis where significant differences across field sites were found between the speakers who had lived in the United States for two to three years. While Durham speakers in this group frequently codeswitched and used quotative frames, the Hickory speakers rarely did so, if at all. There are many possible explanations for why such differences are not significant for those with longer LORS. The difference may reflect the level of tolerance for linguistic variation present within each community. The socio-cultural context in which the Durham participants live may be more open to linguistic innovation, encouraging the use of more varied quotative frames and CS, especially within the first LOR group. This acceptance of linguistic diversity may be a reflection of the community's diversity as represented through the coexistence of AAVE and SWE in the region. In contrast, Hickory's population is ethnically more homogenous, which may lead to tighter community linguistic constraints even within early acquisition stages. Alternatively, it may mean that early variation often levels out to produce regional norms through processes of koineization.

Not all differences present within the first LOR group disappear from the speech of those with longer LORS, however. This may indicate that early variation may lead to established local differences. The patterns of variation observed in the quotative frame analysis support this view. The rich pool of quotative frames found within the speech of the

Durham newcomers also appears in the speech of those who have lived in the United States for over 8 years, so that quotatives like *tell* and *verb* $_+$ *like* cannot be dismissed as mere learner variation that falls away over time. Similarly, CS patterns do not disappear across LOR group, but their type and function are refined. It is entirely possible, although not definitive, that early variation in the quotative and CS systems of learner codes could influence the speech of native English speakers within the speech community.

This study was able to glean such insight through the inclusion of non-native English speakers in the analysis. While such a practice is inappropriate if the researcher approaches the learner varieties as equivalent to native speaker varieties, excluding such data means missing insights regarding the development of dialects as such learner varieties necessarily contribute to the variation within these emerging communities' ecolinguistic niches. If such speakers had not been included, insights regarding the progression of CCR from relatively uniform high levels of consonant cluster reduction to constrained reduction levels would not have been observed. Similar patterns regarding tense and quotative frames, which demonstrate how acquisition levels interact with the ordering of constraints, would have been entirely missed. Observing these processes in action may allow for generalizations that can be applied to future studies of communities with similar heritages.

Understanding variation across LOR allows for a more refined assessment of similarities and differences across field sites, which in turn provides a better basis for comparison with previous research of Latino English in other areas of the country. As the purpose of this paper is to contextualize the variation in an emerging community, looking at differences across LORs and field sites allows for the advancement of this goal. However, this focus on local difference should not obscure the similarities that exist between the field

sites, many of which align with patterns found in other communities as well. Within the syntactic and morphosyntactic variables broader global trends emerge in correspondence with the literature; including differences in CS preferences according to proficiency or interlocutor, and high rates of *be like*, especially to frame tokens of thought. The consistent division of labor between quotative *be like* and *say* is one such finding, where comparisons with global norms demonstrate that the Latino English variety under analysis participates in global trends in syntactic variation. That participants with short LORs favor alternation code switches shows up in data from Dutch-Turkish CS data (Backus 1992) to other studies of Spanish-English codeswitching (e.g. Zentella 1997), indicating that emerging communities can demonstrate structured and consistent variation. Both Durham and Hickory participants demonstrated high levels of consonant cluster reduction, except in bi-morphemic tokens preceding vowels, something that has been noted in the literature on Latino English since Wolfram's (1974) study of Puerto Rican English in New York City. This pattern fits with general findings that prevocalic bimorphemic tokens are the most likely to retain consonant clusters among proficient English speakers. These patterns reveal alignment to the more global trends associated with each variant.

Yet, the results from each analysis illustrate that Latino English in North Carolina is distinct, acclimating in different ways within its new environments. This becomes apparent not only in the less general patterns of the syntactic variants, but is especially clear in the phonetic variants. The wide range of vowel variation provides a key example. While the four participants' vowel charts reveal many commonly-cited Latino English features such as glide weakening on /i/ and /e/ or raised front vowels, the inter-speaker variation is quite diverse, and several of the participants, like David and Jesus, exhibit more extensive

accommodation to Southern English Vernaculars beyond what has previously been observed in Latino English in North Carolina (Wolfram, et al. 2004). That non-pre-nasal /æ/ appears to be lower in Hickory than in Durham, while not statistically significant, could indicate the germs of local vowel production norms where Latino English aligns more with Anglo varieties in Hickory, while African American Vernacular raised /æ/ reinforces Latino English's tendency to have raised front vowels in Durham. The existence of community differences becomes particularly apparent when looking at consonant cluster reduction before a pause in the two locations. As noted in previous studies, Latino English's tendency toward syllable timing may encourage increased reduction of consonant clusters except for in salient positions that carry syntactic information. However, what becomes salient is dependent on community norms. Variation in descriptions of Latino English where some authors found reduction before a pause to be favored (Wolfram 1974) while others find that reduction in such locations is disfavored (Santa Ana 1991) may reflect a difference in acclimation to community norms where founder dialects diverge, rather than a national change in an ethnic minority dialect. Understanding the context of community formation could enhance the predictive power of such findings as features sensitive to local accommodation may be identified.

The introduction to this thesis echoed Blommaert's (2003) call to contextualize analyses of linguistic variables with close observations of demographic trends in order to grant predictive power to sociolinguistic observation. As shown in section 2.3, the field sites under investigation offer very different economic opportunities and have very different demographic make-ups from both each other and previous field sites like Raleigh, NC and Siler City, NC. The impact of this difference is evident in the community differences that do

105

emerge within this study. Some are subtle, such as the use of quotative *tell* in Durham, or the Anglo features found in Marcos's vowel chart. Some fall away quickly, such as CS behavior directed toward community outsiders. The large amount of Southern Vernacular features in David's chart, for instance, could indicate increased contact with Anglos and African American populations. The leveled economic environment where the majority of the town's employment opportunities have been the same for over a century, making Hickory, NC so different from Siler City, NC, strengthens this hypothesis and could explain the differences between certain findings in this study and in previous studies (e.g. Thomas, et al. 2006). Other differences are quite impressive, such as the significantly higher amount of consonant cluster reduction in Durham, which could form out of substrate influences being reinforced by contact with AAVE.

That global variation is locally constrained proves particularly relevant for populations in transit, as seen in the analysis above. What emerges from this process is the understanding that communities exist within regional norms but that community differences fine-tune such norms, giving a local flare to even the most global variant (Buchstaller 2006, 2007). Such truths remain evident even within emerging communities with a heritage language. However, the influence of such a heritage must also be taken into account, as certain features, such as prosody, constrain the types of accommodation that occurs. Understanding how community differences impact dialect variation not only demonstrates that Latino English interacts with majority influence, but also shows that emerging community dialects can be studied and compared to more established ethnic minority dialects in order to understand how such codes reach structured heterogeneity.

References

Angermeyer, Phillipp. 2006. Codeswitching in interpreter-mediated interaction: Comparing four communities in contact. 80th Annual LSA Meeting, Albuquerque.

Backus, Ad. 1992. Patterns of Language Mixing: A Study in Turkish-Dutch Bilingualism, Wiesbaden : O. Harrassowitz.

---2001. The role of semantic specificity in insertional codeswitching: Evidence From Dutch-Turkish. Codeswitching Wordwide II, ed. by R. Jacobson. Berlin: Mouton de Gruyter, 125-154.

Bayley, Robert. 1994. Consonant cluster reduction in Tejano English. Language Variation and Change 3. 303-326.

Bayley, Robert, and Otto Santa Ana. 2004. Chicano English grammar. A Handbook of Varieties of English, Vol. 2: Morphology and Syntax. ed. by Edgar Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton. New York: Mouton de Gruyter. 167-183.

Bell, Allan. 1984. Language style as audience design. Language in Society 13. 145-204.

Bentahalia, Abdelâli; Davies, Eileen. 1992. Code-switching and language dominance. Cognitive processing in bilinguals, ed. by R.J. Harris. Amsterdam: Elsevier, P. 443-58.

Bernstein, Cynthia. 1993. Measuring social causes of phonological variables. American Speech 68. 227-40.

Blyth, Carl Jr., Sigrid Recktenwald, and Jenny Wang. 1990. I'm like, "say what?!": A new quotative in American oral narrative. American Speech 65. 215-227.

Blommaert, Jan. 2003. Commentary: A sociolinguistics of globalization. Journal of

Sociolinguistics 7. 606-623.

Blommaert, Jan, James Collins, and Stef Slembrouck,. 2005. Spaces of multilingualism. Language and Communication 25. 197-216

Brewer, Cynthia A., and Trudy A. Suchan. 2001. Mapping Census 2000: The geography of US diversity. Redlands, CA: Esri Press.

Britain, David. 1991. A geolinguistic analysis of speech variables in the Fens. Colchester: University of Essex.

   --- 2000. The difference that space makes: an evaluation of the application of human geographic thought in sociolinguistic dialectology. Essex research reports in linguistics. 29. 38-82.

Bucholtz, Mary. 1999. Why be normal: Language and identity practice in a community of nerd girls. Language in Society 28. 203-223

Buchstaller, Isabelle. 2001. "He goes and I'm Like: The new quotatives revisited. Paper presented at NWAV 30, North Carolina State University.

   --- 2006. Diagnostics of age-graded linguistic behaviour: The case of the quotative system. Journal of Sociolinguistics 10. 3-30.

   --- 2007. Localized globalization: A multi-local, multivariate investigation of quotative be like. Paper presented at NWAV 36, University of Pennsylvania.

Calvet, Louis-Jean. 2006. Towards on ecology of world languages, trans. by Andrew Brown. Malden, MA: Polity Press.

Carter, Phillip. 2004. The emergence of Hispanic English in the Raleigh community: A sociophonetic analysis. MA thesis. Raleigh, North Carolina: North Carolina State University.

---2007. Phonetic variation and speaker agency: Mexicana identity in a North Carolina middle school. Penn Working Papers in Linguistics: Papers from NWAV 35 13. 1-14.

Chambers, J.K. 2000. World enough and time: Global enclaves of the near-future. American Speech 75. 285-287.

Cheshire, Jenny, and Penelope Gardner-Chloros. 1997.Communicating gender in two languages. Communicating Gender in Contact, ed. by H. Kothoff and R. Wodak Amsterdam: Benjamins. 249-281.

Clyne, Michael. 1987. Constraints on code-switching: how universal are they? Linguistics 25. P. 739-64.

Coulmas, Florian. 2005. Changing language regimes in globalizing environments. International Journal of the Sociology of Language 175/176. 3-15.

Cukor-Avila, Patricia. 2002. She say, she go, she be like: Verbs of quotation over time in African American Vernacular English. American Speech 77. 3-31.

De Anda, Roberto M. 2005. Employment hardship among Mexican-origin women. Hispanic Journal of Behavioral Sciences 27. 43-59.

Deuchar, Margaret, Pieter Muysken, and Sung-Lan Wang. 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. International Journal of Bilingual Education and Bilingualism 10. 298-340.

Downs, Anne, Stevenson Fergus, Yvonne Gagnon, Geeta George, CourtneyGriffiths, and Jamie Newman,. 2000. Chatham County, North Carolina: A Community Diagnosis including Secondary and Qualitative Data Collection and Analysis. Chapel Hill: University of North Carolina at Chapel Hill Dissertation.

Doughty, C. 1991. Second language instruction does make a difference: Evidence from an

empirical study of SL relativization. Studies in Second Language Acquisition 13. 431-469.

Durand, Jorge, Douglass Massey, and Fernando Charvert. 2000. The changing geography of Mexican immigration to the United States: 1910-1996. Social Science Quarterly 81. 1-16.

Durham Public Schools Quick Facts. (2007). http://www.dpsnc.net/

Fasold, Ralph. 1972. Tense marking in Black English: A linguistic and social analysis. Arlingtion: Center for Applied Linguistics.

Ferrara, Kathleen, and Barbara Bell. 1995 Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *Be+like*. American Speech 73. 265-297.

Fought, Carmen. 1999. A majority sound change in a minority community: /u/-fronting in Chicano English. Journal of Sociolinguistics 3. 5-23.

--- 2003. Chicano English in Context. New York: Palgrave MacMillan.

Frazer, Timothy C. 1996. Chicano English and Spanish interference in the Midwestern United States. American Speech *71*, 72-85.

Gal, Susan. 1987. Codeswitching and Consciousness in the European Periphery. American Ethnologist 14. 637-653.

-1988 The Political Economy of Code Choice. Codeswitching: Anthropological and Sociolinguistic Perspectives, ed. by Monica Heller. New York : Mouton de Gruyter. 245-264.

Gardner-Chloros, Penelope. 1995. Code-switching in community, regional and national repertoires: The myth of the discreteness of linguistic systems. One Speaker, Two Languages, ed. by Lesley Milroy and Pieter Muysken. Cambridge, UK:

Cambridge University Press. 68-89.

Gardner-Chloros, Penelope, Reeva Charles, and Jenny Cheshire. 2000. Parallel patterns? A
comparison of monolingual speech and bilingual codeswitching discourse. Journal of
Pragmatics 32. 1305-1341.

Godinez, Manuel, Jr. 1984. Chicano English phonology: Norms vs. interference
phenomena. Form and Function in Chicano English. ed. by Jacob Ornstein-Galicia.
Rowley, MA: Newberry. 42-47.

Godinez, Manuel, Jr. and Ian Maddieson 1985. Vowel Differences between Chicano and
General Californian English? International Journal of the Sociology of Language 53.
43-58.

Gordon, Matthew J. 2000. Phonological correlates of ethnic identity: Evidence of
divergence? American Speech 75. 115-36.

Hartford, Beverley S. 1978. Phonological differences in the English of adolescent female
and male Mexican-Americans. International Journal of the Sociology of Language 17.
55-64.

Haugen, Einar. 1972. The ecology of language, ed. by Anuar Dil, 324-339. Stanford:
Stanford University Press.

Johnstone, Barbara. 2000. The individual voice in language. Annual Review of
Anthropology 29. 405-424.

Kandel, William, and John Cromartie. 2004. New patterns of Hispanic settlement in rural
America. Rural Development Research Report 99. USDA. 1-44.

Kandel, William, and Constance Newman. 2004. Rural Hispanics: Employment and
Residential Trends. Amber Waves 2. 38-45.

Kandel, William, and Emilio A. Parrado. 2005. Restructuring of the US Meat Processing Industry and New Hispanic Migrant Destinations. Population and Development Review 31. 447–471.

Kasarda, John. D., and James. H. Johnson, Jr. 2006. The economic impact of the Hispanic population on the state of North Carolina. Chapel Hill: University of North Carolina Press.

Kerswill, Paul, and Peter Trudgill. 2005. The Birth of New Dialects. Dialect Change, ed. by Peter Auer, Frans Hinskens, and Paul Kerswill. 196- 220. Cambridge, UK: Cambridge University Press.

Kohler, Adiana, and Melissa Lazarín. 2007. Hispanic Education in the United States. National Council of La Raza Statistical Brief 8: 1-16.

Labov, William. 1963. The social motivation of a sound change. *Word*, 19: 273-309.

--- 1966. The social stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics.

--- 1989. The child as linguistic historian. Language Variation and Change 1. 85-97.

Labov, William, Paul Cohen, Clarence Robins, and John Lewis. 1968. A study of the non-standard English of Negro and Puerto Rican speakers in New York City. United States Office of Education Final Report, Research Project 3288.

Liaw, Kao-Lee, and William H. Frey, 2007. Multivariate explanation of the 1985-1990 and 1995-2000 destination choices of newly arrived immigrants in the United States: The beginning of a new trend? Population, Space and Place 13. 377-399.

Lowell, Lindsey, & Roberto Suro. 2002. The improving educational profile of Latino immigrants. Pew Hispanic Center Report. Washington, D.C.: Pew Charitable Trusts and the USC Annenberg School for Communication.

Massey, Douglass. 1990. The social and economic origins of immigration. Annals of the American Academy of Political and Social Science 510. 60-72.

Massey, Douglass, Jorge Durand, and Nolan Malone. 2002. Beyond smoke and mirrors: Mexican immigration in an era of economic integration. New York, NY; Russell Sage.

McClure, Erica. 1997. The Interrelationship between form and function in written national language – English codeswitching: Evidence from Mexico, Spain, and Bulgaria. Codeswitching Worldwid, ed. by R. Jacobson. The Hague: Mouton de Gruyter. P. 125-150.

Mesthrie, Rajend. 1993. Koineization in the Bhojpui-Hindi diaspora. International Journal of the Sociology of Language 99. 25-44.

Milroy, Lesley; and Li Wei. 1995. A social network approach to code-switching: the example of a bilingual community in Britain. One Speaker, Two Languages, ed. by Lesley Milroy and Pieter Muysken. Cambridge, UK: Cambridge University Press. 136-57.

Mohney, Kirk F., and Laura A.W. Phillips. 1988. From Tavern to Town. Hickory, NC: Hickory Landmarks Society.

Moriello, Becky. 2003. 'I'm feeksin' to move . . . ': Hispanic English in Siler City, North Carolina. MA thesis. Raleigh, North Carolina: North Carolina State University.

Mufwene, Salikoko S. 2001. The ecology of language evolution. Cambridge, UK: Cambridge University Press.

Mühlhäusler, Peter. 1996. Linguistic ecology: language change and linguistic imperialism in
  the Pacific region. London: Routledge.

Muysken, Pieter. 2000. Bilingual Speech: A Typology of Code-Mixing. Cambridge, UK:
  Cambridge University Press.

Myers-Scotton, Carol. 2002. Contact linguistics: Bilingual encounters and grammatical
  outcomes. Oxford: Oxford University Press.

  ---1993a. Social Motivations for Codeswitching: Evidence from Africa.

  Oxford: Oxford university Press.

  --- 1993b. Common and uncommon ground: Social and structural factors in

  codeswitching. Language in Society 22. 475-503.

Netstate. 2007. http://www.netstate.com/states/maps/images/nc_outlinecap.gif

North Carolina Profile. 2002. http://eslmi23.esc.state.nc.us/ncp/genInfo/industry.aspx

Paolillo, John. 2002. Analyzing linguistic variation: Statistical models and methods.
  Stanford: CSLI Publications.

Penfield, Joyce, and Jacob Ornstein-Galicia. 1985. Chicano English: An ethnic contact
  dialect. Varieties of English Around the World, General Series, 7.

Poplack, Shana. 1978. Dialect acquisition among Puerto Rican bilinguals. Language in
  Society 7. 89-103.

  --- 1980. Sometimes I'll start a sentence in Spanish *y termino en español.* Linguistics
  18. 581-618.

  --- 1988. Language status and language accommodation along a linguistic boarder.
  Language Spread and Language Policy, ed. by Peter Lowenberg. Washington, DC:
  Georgetown University Press. 90-118.

Poplack, Shana, Susan Wheeler, and Anneli Westwood. 1987. Distinguising language-
  contact phenomena: evidence from Finnish-English bilingualism. The Nordic
  Languages and Modern Linguistics 6, Proceedings o the Sixth International
  Conference of Nordic and General Linguistics in Helsinki. ed. by Pirkko Lilius
  and Mirja Saari.

Rickford, John, Thomas Wasow, Arnold Zwicky, and Isabelle Buchstaller. 2007.. Intensive
  and quotative *all*: Something old, something new. American Speech, 82. 3-31.

Roeder, Rebecca V. 2006. Ethnicity and sound change: Mexican American
  accommodation to the Northern Cities Shift in Lansing, Michigan. Ph.D.
  dissertation, Michigan State University.

Romaine, Suzanne, and Deborah Lange. 1991. The use of like as a marker of reported speech
  and thought: A case of grammaticalization in progress. American Speech 66. 227-
  279.

Sanchez, Tara, and Anne Charity. 1999. Use of be like and other verbs of quotation in a
  predominantly African-American community. Paper presented at NWAV 28,
  Toronto.

Santa Ana, Otto. 1991 Phonetic simplification processes in the English of the barrio: A
  cross-generational sociolinguistic study of the Chicanos of Los Angeles.
  Doctoral dissertation, University of Pennsylvania.
  --- 1993. Chicano English and the Nature of the Chicano Language Setting.
  Hispanic Journal of Behavioral Science 15. 3-35.
  --- 1996. Sonority and syllable structure in Chicano English. Language Variation
  and Change 8. 63-91.

Sawyer, Janet B. 1959. Aloofness from Spanish influence in Texas English. *Word* 15. 270-81.

Singler, John Victor. 2001. Why you can't do a VARBRUL study of quotative frames and what such a study can show us. University of Pennsylvania Working Papers in Linguistics 7. 257-278

Soja, Edward. 1989. Postmodern geographies: The reassertion of space in critical social theory. New York: Verso.

Syrdal, Anne K. and H. S Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America 79.1086-1100.

Tagliamonte, Sally, and Rachel Hudson. 1999 *Be like* et al. beyond America: The quotative system in British and Canadian youth. Journal of Sociolinguistics 3, 147-172.

Tannen, Deborah. 1989. Talking voices: Repetition, dialogue, and imagery in conversational discourse. Cambridge, UK: Cambridge University Press.

Thomas, Erik; Phillip Carter, Elizabeth Cogshall. 2006. Acoustic evidence for new dialect formation. Paper presented at the American Dialect Society Annual Meeting: Albuquerque.

Thomas, Erik. 1989. The implications of /o/ fronting in Wilmington, North Carolina. American Speech 64. 223-32.

---1993. Why we need descriptive Studies: Phonological Variables in Hispanic English. Proceedings of the First Annual Symposium about Language and Society *(SALSA I)*, ed. by Robin Queen and Rusty Barrett. Austin: Department of Linguistics, University of Texas at Austin. 42-49.

--- 1997. A rural/metropolitan split in the speech of Texas Anglos.

 Language Variation and Change 9. 309-332.

--- (2001). An Acoustic Analysis of Vowel Variation in New World English.

Durham: Duke University Press.

Thompson, Roger M. 1975. Mexican-American English: Social correlates of regional

pronunciation. American Speech 50. 18-24.

Treffers-Daller, Jeanine. 1997. Variability in code-switching styles: Turkish-German

 code-switching patterns. Codeswitching Worldwide, ed. by R. Jacobson. The Hague:

Mouton de Gruyter.177-198.

 --- 1992. French-Dutch codeswitching in Brussels: Social factors explaining its

disappearance. Journal of Multilingual and Multicultural Development 13. 143-156.

Trudgill, Peter, Elizabeth Gordon, Gillian Lewis, and Margaret Maclagan. 2000.

Determinism in new dialect formation and the genesis of New Zealand English.

Journal of Linguistics 36. 299-318.

Trudgill, Peter. 1986. Dialects in Contact. Languages in Society 10, ed. by Peter Trudgill.

Oxford, Great Britain: Basil Blackwell, ltd.

Veatch, Thomas. 1991. English vowels: Their surface phonology and phonetic

implementation in vernacular dialects. Doctoral dissertation,

University of Pennsylvania.

Wald, Benji. 1984. The status of Chicano English as a dialect of American English. Form

and Function in Chicano English, ed. by Jacob Ornstein-Galicia and Allan

Metcalf, 14-31. Rowley, MA: Newberry House.

Wolfram, Walt. 1969. A sociolinguistic description of Detroit Negro speech. Arlington:

Center for Applied Linguistics.

---1974. Sociolinguistic aspects of assimilation: Puerto Rican English in New York City. Arlington, VA: Center for Applied Linguistics.

Wolfram, Walt, Becky Childs, and Benjamin Torbert. 2000. Tracing language history through consonant cluster reduction: Comparative evidence from isolated dialects. Southern Journal of Linguistics 24. 17-40.

Wolfram, Walt, Phillip Carter, and Becky Moriello. 2004. Emerging Hispanic English: New dialect formation in the American South. Journal of Sociolinguistics 8. 339-358.

Wolfram, Walt, and Natalie Schilling-Estes. 1998. American English: Dialects and variation. Malden, MA: Blackwell.

Wolfson, Nessa. 1982. On tense alternation and the need for analysis of native speaker usage in second language acquisition. Language Learning 32. 53-68.

United States Census Bureau (2007). http://factfinder.census.gov

U.S. Census Bureau, Census 1990, 2000 Summary File 1, Matrices P1, P8.

Yule, George, and Terrie Mathis. 1992. The role of staging and constructed dialogue in establishing speaker's topic. Linguistics 30. 199-215.

Zentella, Ana Celia. 1997. Growing Up Bilingual: Puerto Rican Children in New York. Malden, MA: Blackwell.