Emerging Hispanic English in the Southeast U.S.: Grammatical Variation in a Triethnic Community

by

Erin Elizabeth Callahan-Price

Department of English
Duke University

Date: _________________________

Approved:

___________________________
Walt Wolfram, Supervisor

___________________________
Erik Thomas

___________________________
Robin Dodsworth

___________________________
Dominika Baran

___________________________
Julie Tetel Andresen

___________________________
Edna Andrews

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of English in the Graduate School of Duke University

2013
ABSTRACT

Emerging Hispanic English in the Southeast US: Grammatical variation in a triethnic community

by

Erin Elizabeth Callahan-Price

Department of English
Duke University

Date: ________________

Approved:

___________________________
Walt Wolfram, Supervisor
___________________________
Erik Thomas
___________________________
Robin Dodsworth
___________________________
Dominika Baran
___________________________
Julie Tetel Andresen
___________________________
Edna Andrews

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of English in the Graduate School of Duke University

2013
Abstract

This study investigates variable past tense marking patterns in an emerging variety of N.C. Hispanic English (n=44) spoken by language learners at three Length of Residency (LOR) groups in three schools in Durham, NC in terms of 1. lexical semantics (Andersen & Shirai 1996, Bayley 1999), 2. frequency (Guy & Erker 2012) 3. discourse structure (Kumpf 1984, Bardovi-Harlig 1998) and 4. verb class and phonological environment (Wolfram 1985, Bayley 1994). Statistical results show significant effects of verb class, lexical aspect, and frequency and interacting effects of verb class and frequency (specifically, suppletives like copula are simultaneously highly frequent and highly phonetically salient). A subsample coded for the discourse factor shows some evidence for the correlation of copula and backgrounding function. An analysis of consonant cluster reduction patterns (CCR) demonstrates dialect acquisition of variable constraints (e.g. in terms of N.C. AAVE), namely phonological environment (\( _P > _C > _V \)) and morphemic status (monomorpheme > bimorpheme), though the significant result for morphemic status is ultimately shown to be due to collinearity with phonological environment (i.e. bimorphemic cluster/regular past tense –ed verbs occur in prevocalic contexts) in both N.C. HE and AAVE. Pedagogical applications are discussed, including accurately identifying English Language Learners (ELLs) in the context of local/regional accommodation.
Dedication

Dedicated to my wife, Alissa Jo Callahan-Price, who leaves me speechless.
Contents

Abstract ................................................................................................................................. iv

List of Tables ........................................................................................................................ x

List of Figures ....................................................................................................................... xi

Acknowledgements ........................................................................................................... xiii

1. Introduction ...................................................................................................................... 1
   1.1 Overview of the Study: The language of ‘Marcos’ ...................................................... 1
   1.2 Study design: Past tense unmarking as a sociolinguistic variable ......................... 3

2. Review of Literature ....................................................................................................... 7
   2.1 Why study interlanguage? Implications for variationism & SLA ......................... 7
   2.2 Tense unmarking: The variationist rubric ................................................................. 19
      2.2.1 Wolfram et al.’s Vietnamese English studies (1980s) ...................................... 19
      2.2.2 The Chinese English studies (1990s) ............................................................... 22
   2.3 Tense unmarking studies: The discourse rubric ...................................................... 26
      2.3.1 Foreground vs. background .............................................................................. 26
      2.3.2 Function of the historical present in narratives .............................................. 28
   2.4 Tense unmarking studies: The role of (two kinds of) aspect .................................. 29
      2.4.1 Semantic/lexical constraints: Akionsart ........................................................... 30
2.4.2 Grammatical constraints: Point-of-View aspect ......................... 31

2.5 Next steps .................................................................................. 36

3. Traditions in variationist sociolinguistics ...................................... 39

3.1 The sociolinguistic variable ......................................................... 39

3.2 Case study 1: Copula deletion in early AAE ............................... 41

3.3 Case Study 2: CCR across English varieties .............................. 45

3.4 CCR as a roadmap for evolving variationist methodology .......... 49

3.4.1 Classical variable rules ......................................................... 49

3.4.2 VARBRUL .............................................................................. 51

3.4.3 Quantitative analysis Post-VARBRUL ................................. 52

3.5 The speech community: Intersections and innovations ............. 55

3.5.1 Ethnolects ............................................................................. 56

3.5.2 Conclusions ........................................................................... 60

4. Site, Speakers and Methods .......................................................... 62

4.1 Durham speech community: demographics and trends ............. 62

4.2 Ethnographic evidence ............................................................... 65

4.3 Speakers ...................................................................................... 67

4.4 Data and coding procedures ....................................................... 69
4.4.1 Overview of coding methods for unmarked tense variable ..........69
4.4.2 Variationist factors ...........................................................................70
4.4.3 Lexical/inherent aspect ....................................................................72
4.4.4 Discourse structure ........................................................................73
4.4.5 Other exclusions .............................................................................78
4.4.6 Statistical methods ..........................................................................78

5. Results ..................................................................................................81

5.1 Introduction ..........................................................................................81
5.2 Overall distribution of tokens ..............................................................82
5.3 Results by linguistic factor group ........................................................83
  5.3.1 Interpretation of logistic regression statistics ..................................84
  5.3.2 Phonological environment ..............................................................87
  5.3.3 Lexical Aspect ................................................................................89
  5.3.4 Verb class .....................................................................................92
  5.3.5 Frequency .....................................................................................98
5.4 Results by discourse factor group ........................................................103
5.5 Results by social factor group ..............................................................107
5.6 Summary of results ............................................................................110
6. HE in the community context ................................................................. 112

6.1 HE and AAE in contact ................................................................. 113

6.1.1 AAE contact features ............................................................. 113

6.1.2 CCR in HE and AAE .............................................................. 116

7. Conclusions and Applications ............................................................ 127

7.1 Review of general findings ........................................................... 127

7.2 Institutional practice: Identifying English Language Learners .......... 128

7.3 Future directions ........................................................................... 133

References .......................................................................................... 135

Biography ............................................................................................ 148
List of Tables

Table 1: Semantic Features for Vendler’s (1967) four event classes ................................................. 32

Table 2: Inherent Aspect with Past and Progressive Inflections in mother’s speech (reprinted from Shirai and Andersen 1995: 751) ................................................................. 34

Table 3: Preterit tense by Aspectual Class in Mexican-Origin Children’s Spanish Narratives (adapted from Schecter and Bayley 2002: 125) ................................................................. 36

Table 4: Variables in quantitative analysis of unmarked tense ............................................................ 69

Table 5: Logistic regression results for linguistic (internal) factors ..................................................... 84

Table 6: AAE Features in Durham HE .................................................................................................. 113

Table 7: Factor weights for 3 CCR factor groups .............................................................................. 117

Table 8: Oral IPT Test by skill area (syntax) with HE examples .......................................................... 131
List of Figures

Figure 1: Incidence of unmarked tense (Unm) for types of irregular verbs, by length of residency (LOR). (Reprinted from Wolfram and Hatfield, 1986, p. 21.) ........................................... 21

Figure 2: Sample of narrative text from Adamson (2009: 64) coded for foreground vs. background ........................................................................................................ 27

Figure 3: The relationship between perfect and past (reprinted from Andersen and Shirai 1994: 149) ........................................................................................................... 33

Figure 4: Percentage of copula deletion by age and structural environment .......... 42

Figure 5: Alternative trajectories of change over the early lifespan: (a) ‘Roller coaster’ trajectory; and (b) Curvilinear trajectory (from Van Hofwegan and Wolfram 2011: 439) 44

Figure 6: Hispanic Population Trends in the 1990s and early 2000s ......................... 63

Figure 7: Race for Durham (city), N.C. (%) ................................................................ 64

Figure 8: Location of 3 fieldwork sites in Durham, N.C. ........................................... 66

Figure 9: Percentage unmarked tokens by speaker ................................................. 83

Figure 10: Raw counts of –t/-d absence by phonological environment .................... 88

Figure 11: Raw Counts of surface past tense marking by lexical aspect .................... 89

Figure 12: Percentage of past tense marking by lexical aspect (copula excluded) .... 92

Figure 13: Raw counts of past tense marking by verb class ..................................... 93

Figure 14: Percentage of past tense marking by verb class ..................................... 94

Figure 15: Percentage of unmarking by verb class and proficiency for 3 studies of English interlanguage ................................................................................. 95

Figure 16: Token by raw frequency with reference line for high vs. low .................... 100

Figure 17: Percentage of marking by frequency and LOR ........................................ 101

Figure 18: Verb class by Frequency and LOR .......................................................... 102

Figure 19: Percentage of past tense marking by speaker and narrative function (foreground vs. background) .......................................................... 104
Figure 20: Background forms and marking ................................................................. 105
Figure 21: Foreground forms and marking ............................................................... 106
Figure 22: Percentage of past tense marking by speaker LOR ................................... 109
Figure 23: Factor weights for CCR by LOR .............................................................. 117
Figure 24: Factor weights for CCR by phonological environment for Durham HE and nine English varieties ................................................................. 119
Figure 25: Percent CCR by phonological environment for Durham HE and six English varieties ........................................................................................................ 120
Figure 26: Factor weights by morphemic status for CCR for Durham HE and six English varieties ........................................................................................................ 122
Figure 27: Percent CCR by morphemic status for Durham HE and five English varieties ........................................................................................................ 122
Figure 28: Mean Frequency for all tokens (marked and unmarked) by morphemic status and phonological environment: Durham AAVE & Durham HE ......................................... 125
Figure 29: Inference Tree for Durham HE CCR .......................................................... 126
Acknowledgements

My dissertation committee is made up of, as my wife would say, some of the most solid people I know. Robin Dodsworth has, very simply, taught me what it means to be a woman in academia. She is one of the hardest-working, most selfless teachers I have ever met; she encouraged me when I didn’t realize I needed it and was tough on me when I thought I didn’t need it. Erik Thomas took a chance on me and invited me to come do fieldwork in south Texas with him back in 2006 when he barely knew me. He is one of the smartest and most sincere people I know. Dominika Baran has been a consistent source of support and advice, my everyday touchstone and motivator at Duke, and now, a friend. Julie Tetel and Edna Andrews have advocated for me, both administratively and intellectually, since I came to Duke. Mary Kohn and I have been in the trenches together too many times to count; here’s no one else I’d want in my corner, academically or personally. Lori Fernald Khamala has provided me with perspective when I needed it. Maryscot Mullins has provided impeccable administrative support that is always underlined by kindness. Thanks also to Sashi Rayasam, Alice Anne Kern, Mary Holt and Daniel Isenberg in DPS schools for opening up their classrooms to a bunch of graduate students with recorders. Even more thanks to the kids at E.K. Powe, Rogers-Herr, and Chewning Middle who made this dissertation possible.

Finally: anyone who knows him will tell you that Walt Wolfram will make you a stellar sociolinguist. I could never have predicted the extent to which he would make me a better person.
1. Introduction

Marcos: I’m more of American than Mexican. I like to eat like, Chinese, Biscuitville, and all that.

Interviewer: Mm-hm. ¿Y um, que — por qué sientes más americano que mexicano? (‘And um, what — why do you feel more American than Mexican?’)

Marcos: Because I know more… more people recognize me as American than Mexican. And like, I know more English. I know how to read it, speak it, and everything. I don’t even know the ABCs in Spanish. That’s why.

Int: Pero… ¿tienes ganas de aprender español? (‘But… do you want to learn Spanish?’)

Marcos: No.

Int: ¿Por qué no? (‘Why not’?)

Marcos: I just can’t. I already try [sic] but I can’t. I can’t even read the Bible.

1.1 Overview of the Study: The language of ‘Marcos’

“Marcos” (pseudonym) is a 13-year old middle school student living in the medium-sized southern city of Durham, NC. He came with his family from the rural state of Michoacán, Mexico, to the United States when he was three years old. His parents, undocumented immigrants, work as a cook and a house painter in Durham. His favorite hip hop artist is Lil’ Wayne. While he claims he “can’t learn” Spanish, Marcos later responds to the interviewer’s questions in completely fluent spoken Spanish and even scores fairly well on a standardized Spanish academic achievement test, the Woodcock-Johnson (WJ-III) (Schrank et. al 2005). Marcos himself says he “doesn’t even
know the ABCs in Spanish” — but he is officially classified by the Durham Public School system as an English Language Learner.

The language of speakers like Marcos is representative of the multifaceted and sometimes self-contradictory set of questions surrounding language shift and identity in the mid-Atlantic South at the turn of the twenty-first century. What kind of English has Marcos learned — and what kind will he end up speaking? The Southern White Vernacular English (SWVE) of his middle school principal, a middle-aged, male, white Durham native? The African American English (AAE) of a majority of his classmates? Or the ‘Spanish-sounding’ English of his best friend’s younger siblings, some of whom can understand Spanish but can’t speak it at all?

This dissertation focuses on the study of language variation and change in southeastern U.S. varieties of Hispanic English (HE): in short, the language of speakers like Marcos. Since the radical increase in in-migration to North Carolina by Hispanic immigrants — 394% between 1990 and 2000, and 111% between 2000 and 2010 — students from Spanish-speaking countries account for over half of the total enrollment growth in North Carolina Public Schools (Kasarda & Johnson 2006). Though Hispanic (English-speaking) communities have populated the southwest and western U.S. for generations, the population explosion in the southeastern United States gives researchers a unique

---

1 The term ‘Hispanic’ (vs. ‘Latino,’ ‘Chicano,’ etc.) was chosen because it was the term most frequently used by speakers (in Durham and Granville County, N.C. Schools, during 2002-2005, where I worked as a K-12 ESL teacher) to describe themselves.
window into the first stages of ethnic dialect formation—beginning with second dialect learning and culminating in the construction of a distinct socioethnic variety.

As such, this study examines patterns of English language-learner variation in the speech of a group of HE speakers, in one elementary school and two middle schools, all in Durham, NC. It examines the sociolinguistic variable of past tense unmarking—the occurrence of verb forms which are formally unmarked in past tense contexts where standard varieties of English use simple past forms (e.g. Marcos’s response, “I already try that” vs. Standard English “I already tried that”). By examining trends in this variable unmarking of tense, the study tests 1) the extent to which English language learners are accommodating to the regional and ethnic (AAE and SWVE) varieties of their peers, and 2) which internal/universal factors guide this sociolinguistic process of dialect acquisition, specifically, a) the verb’s sound pattern b) its inherent meaning, or c) the frequency of the verb in the discourse.

1.2 Study design: Past tense unmarking as a sociolinguistic variable

The task of a variationist sociolinguistic study is to describe and explain the systematic variation of language structures. Thus, a sociolinguist might study how forms of the verb to be are variably deleted in AAE (“He cute” vs. “She’s a linguist”) depending on subject type (full noun vs. pronoun), phonetic environment (preceding consonant vs. preceding vowel), and following grammatical environment (adjective vs. verb). As such, every variationist study must begin with a linguistic variable—a specific, quantifiable
language structure which seems to act variably in different environments (as with the behavior of the verb to be in AAE). This study selects unmarked past tense forms (e.g. “I already try that”) as a variable which consistently appears in the speech of newly arrived immigrants in Durham, N.C. Though the acquisition of past tense has been well-documented in the SLA literature (Andersen & Shirai 1996, Bardovi-Harlig 1998), there have not yet been many variationist takes on the process from a sociolinguistic point of view. For this same reason— we have a good idea of how the process may work from a language acquisition standpoint, but not from a social one— the structure is an optimal site for investigating how the process operates sociolinguistically. Do language learners in Durham show the same kind of universal effects documented for second language learners of English in general— specifically more past tense marking in the aspectual classes of so-called and ‘achievement’ and ‘accomplishment’ verbs (Vendler 1967)— more marking in sentences like She broke the machine vs. When I was growing up, I live in Monterrey? At the same time, if speakers have contact with speakers of African American English (AAE) and Southern White Vernacular Englishes (SWVE), what effect do the phonological (sound structure) patterns in those varieties have on how speakers develop and produce past tense forms? Specifically, do newly emerging Hispanic English speakers acquire the same kind of consonant cluster reduction patterns documented for these varieties in the variationist literature, where, for example, in AAE, the final [t] of a noun-like form like ‘test’ gets deleted more often than the [t] in a verb-like form like ‘picked,’ or more often before another consonant than a vowel or pause (‘picked
strawberries’ vs. ‘picked over’ or ‘the girl that got picked’)? Furthermore, what can these
linguistic processes tell us about how speakers negotiate their newly-forming identities
which respect to social categories like language proficiency, age, and gender?

Within the context of an emerging ethnolect, a comprehensive exploration of past
tense unmarking is especially valuable since the same structure varies in quantifiable
ways across multiple linguistic levels—phonology (one regular English tense marker is
formed in consonant clusters which are variably reduced according to well-known
linguistic constraints), morphology (irregular verbs may be variably unmarked for
tense), discourse function (unmarked verbs may have a specific purpose in narratives),
language development (increased LOR may affect overall rates), and social function
(unmarking and/or cluster reduction may highlight group membership categories in the
speech community). In this way, past tense unmarking illustrates both a cross-section of
variable processes (i.e. in order to understand how language varies across structural
levels) but, in a larger sense, also demonstrates how socially- and linguistically-situated
meanings themselves emerge in the context of the language learner’s ‘ethnolinguistic
repertoire’ (Benor 2010), i.e. in order to understand how language itself works. What
linguistic forms and functions (e.g. phonological vs. narrative variants) stabilize as
language learning informs a dynamic process of dialect formation (i.e. as a group’s
unique, ethnically-indexed identity forms in a new community)? Do native- and non-
native speakers show the same kinds of same linguistic and social constraints (e.g. is
tense unmarking a nonnative effect)? How do these distinctive features enter (and exit)
the ethnolinguistic repertoire/grammar over time (i.e. in terms of LOR)? If they do enter the repertoire, how do (or do they) they demarcate to both insiders and outsiders which speakers are members of the group?

Data for the project come from 2006-2008 fieldwork funded as part of a National Science Foundation Research Project (BCS-054139) on emerging Hispanic English in North Carolina. I was conducted fieldwork in the three field sites in Durham, NC: E.K. Powe Elementary, Rogers-Herr Middle School, and Chewning Middle School. The fieldwork culminated in a collection of 128 bilingual interviews (from a half-hour to an hour and a half in length) with speakers from the Durham site, ranging in age from 3rd grade to 8th grade. A subsample of 42 English language interviews were ultimately coded in in terms of a variety linguistic and social factors, including verb class (e.g. irregular suppletives like go/went vs. regular pick/picked), phonological environment (following vowel vs. consonant), lexical aspect (achievement vs. activity), as well as speaker’s Length of Residency (LOR), gang affiliation, literacy in English vs. Spanish, age, gender, and country of origin. Data were then analyzed using statistical programs, including SPSS, and results are presented in graphic form as well as textually. A concluding chapter summarizes trends and suggests areas for future research as well as pedagogical implications for English as a Second/Foreign Language (ESL/EFL) instructors.
2. Review of Literature

2.1 Why study interlanguage? Implications for variationism & SLA

Since the publication of Dickerson’s 1975 article, “The learner’s interlanguage as a system of variable rules,” a growing contingent of sociolinguists, trained to analyze (L1) linguistic variation using various quantitative methods, have worked from Selinker’s (1972) INTERLANGUAGE model to analyze the speech of second language learners. These studies have ranged from Dickerson’s examination of the variable productions of /t/ for a group of Japanese English Language Learners (ELLs) (Dickerson 1975, 1976) to Bayley’s (1991, 1994) study of constraints on the -t/d variable in Chinese learners of English to Tarone’s work (1979, 1981) on attention and style-shifting in interlanguage systems. These studies have constituted a small but promising foundation for the emerging field of variation in the speech of second language learners.

In providing a rationale for the existence of this new line of inquiry, Preston (1989: 2-3) writes:

Since there are probably more bi- and multi-linguals than monolinguals in the world, it should be an idea especially abhorrent to sociolinguists that their special interest can be pursued adequately in ignorance of the messy data produced by such speakers. Many who are known as sociolinguists prefer to be called linguists, assuming that their perspective on language in its broader social context is necessary to any complete understanding not only of the interactional functions but also of the internal make-up of language systems…SLA provides another perspective from which language and its structure may be investigated.

As a unique kind of speech community, the speech of second (or third-) language learners should be a natural object of inquiry for sociolinguists, as well as the kind(s) of
competence the knowledge of this speech community represents. The ‘messy data’

Preston refers to bears more than a passing resemblance to Labov’s (1969) earliest data
on language variation in the copula, where the systematicity of an individual’s speech,
unless positioned in the context of a larger speech community, resembled
“unaccountable and sporadic variation” (1969: 759). In providing an empirical rationale
for the study of variable rules, Labov was, in part, appealing for greater cooperation
from the generative community to pursue unified solutions for understanding the
nature of variation. To that end, he writes:

More generally, the paper is directed at the methodological problem which
seems to me of overriding importance in linguistics at the moment: to connect
theoretical questions with a large body of inter-subjective evidence which can
provide decisive answers to those questions (757).

A conversation between SLA researchers and sociolinguists would constitute the same
kind of methodological advance Labov argues for. While SLA research has typically
considered interlanguage variation in terms of a categorical, intuited choice between
two forms (frequently, more or less target-like), sociolinguists are poised to provide
rigorous descriptions of the principled nature of the variation between these forms.

In addition to providing sociolinguists with a unique type of data, second-
language learners may ultimately help sociolinguists unify their accounts of synchronic
language variation and language change. To the degree that second language acquisition
involves the internal development of linguistic systems in the context of the social life—
a process manifested in misperceptions, re-balancings, re-evaluations, regularizations,
innovations, and standard-settings—it may bear fruit for our investigations of socially- and linguistically-conditioned language change.

To give an example: it was of special importance to sociolinguists to demonstrate, using the construct of APPARENT TIME, that historical processes of language change could be reliably observed in real-time. Consequently, sociolinguists are now especially well-versed in ‘the use of the present to explain the past’ (Labov 1963). It is intriguing, then, that the second-language acquisition data, (e.g. in the form of implicational arrays) has been characterized by some sociolinguists as representing a form of ‘violent language change’ (Major 1994a; see also Preston 1989: 32-33). In examining how interlanguage data (especially contact data) ‘select’ which linguistic tendencies survive into a full-fledged ‘ethnolinguistic repertoire’ (Benor 2010) we are perhaps at another crossroads where two perspectives may cross in order to address fundamental questions like “What is the most general form of a linguistic rule?” and “How do systems of rules change and evolve?” (Labov 1994: 760, cf. Weinreich, Labov, and Herzog 1968). ¹ In other words, to what degree can an analysis of longitudinal development (in the individual) explicate or model processes of language change (in the speech community)? Or, vice-versa, to what degree can variationist constructs designed to explain language change (e.g. real vs. apparent time) account for successive stages in interlanguage development?

¹ Labov (1994: 760) also pointedly asks (in the same series of questions) “How do languages, originally diverse, combine within a bilingual speech community?”
This study considers data which come from a privileged window of linguistic time: the stretch between the ‘psycholinguistic’ process of second language acquisition and the ‘sociolinguistic’ process of dialect emergence. In exploring how a range of constraining factors—including subtle linguistic and cross-linguistic facts, discourse-level/interactional goals, and even developmental universal processes—interact with a single variable, \textsc{past-tense unmarking}\footnote{I originally chose the term ‘unmarking’ (vs. ‘lack of past tense marking,’ etc.) to remain consistent with Wolfram and Christian’s Vietnamese English studies, one of the first to examine past tense marking in interlanguage. I do not mean to suggest that speakers are removing marking in some way, only that (in a descriptive sense) the verb forms are not overtly marked for Standard English type past tense.}, in the speech of second-language learners who are also emerging dialect speakers—this study hopes to join the conversations of sociolinguists working on SLA variation, as well as a larger conversation with general linguistic theory.

Wolfram (1991: 105) frames fundamental questions which may need to be answered before any meaningful expedition sets off to chart interlanguage variability:

1. What is the nature of interlanguage variability, both systematic and unsystematic?
2. Beyond simple (correlational) descriptions, can we provide a unitary model which will link description and explanation?

Is all interlanguage variation systematic? Preston (1993) points out that an exhaustive search of factors on variation in a particular environment (linguistic or otherwise)\footnote{Preston’s (1989: 194-238) the impressive taxonomy includes ‘fifty-some-odd’ factors of ‘sociolinguistic concern,’ including linguistic, individual, interactive, and sociological processes relevant to the study of language variation.} must be carried out before variation can be deemed entirely unsystematic; he is wary of
accounts of so-called ‘free variation’ and writes “I am suspicious that language variation which is to be influenced by nothing at all is a chimera…” (155; see also 2009:99).

Accordingly, there are robust findings in the literature where reanalyses have led to more focused understandings of apparently ‘free’ variation. Schachter (1989), for example, reanalyzes the variable negation patterns in the speech of a speaker in Cazden et al. (1975): Jorge, a 12-year old Spanish-speaking student from Colombia. In a longitudinal six-week study which began one month after Jorge entered the U.S., Schachter finds that both syntactic development (from undifferentiated Eng. no + N/V/PP/etc. to elaborated Eng. no + N, not + V…) and functional category (rejection, denial, correction, etc.) systematically constrain Jorge’s negation strategies. Schachter emphasizes that the complexity of decisions and procedures of the analyst must match the complexity of the learner’s patterns of variation so that “it is the variation itself that should serve as a challenge to further and deeper analysis” (p. 131). Careful analyses like Schacter’s provide a rationale for prevailing variationist opinion that ‘free variation’ usually is not completely free.

On the dissenting side, Ellis (1994, 1999; see also 1985, 1989 for earlier lists) defines free variation as that which may not be accounted for by a list of five factors:

1. the situational context, covering such factors as the setting and the addressee
2. the illocutionary force of an utterance
3. the linguistic context
4. the discourse context (i.e. the rhetorical mode)
5. the planning conditions under which performance takes place
Ellis argues that the remaining free variation is a product of ‘item’ vs. ‘system-learning’: acquisition occurs in the context of “loose lexical networks” which are the product of implicit learning: in this context, free variation occurs “when learners add items to those they have already acquired and before they analyse these items and organize them into a system” (1990: 460).

Along these lines, Young (1996) describes free variation as an early stage in the acquisition of English articles by native speakers (NS) of Czech and Slovak learners of English. Here, formal strategies for encoding semantic and discourse information related to reference and topic continuity, which are encoded by different forms in the learners’ L1s (for example, via verbal aspect or case marking), are shown to vary according to progressing stages in the interlanguage system. Definite articles show a distinct pattern at the lower proficiency levels, however: they are over-generalized with no clear form-function relation (at least none of those coded in the study). Young notes that this result tallies with similar findings by Huebner (1983) and Chaudron and Parker (1990). Both studies report the same “nonsystematic flooding” of definite articles in the low proficiency stage of their subject(s): an adult, Hmong-speaking Laotian refugee named Ge, and a group of 40 L1 Japanese ELLs. These results lead Young to conclude that free variation of an L2 form may occur under specific conditions:

1. The L2 form does not have a corresponding form in the L1
2. The L2 form is perceptually salient
3. There is no clear form-function relation between the L2 form and meaning.\

---

† Presumably, this set of factors would apply only to cases of free variation of elements in the L2 morphosyntactic system (cf. Lavandera 1978)
4. Free variation in the L2 form consists of initial overuse of the form.
5. Systematic use of the form begins when the form disappears from some environments. (1996: 170)

Preston (1989, 1996, 2009) is especially candid in criticizing SLA studies which relegate so-called ‘sociolinguistic’ concerns to only those factors usually within, for example, the ethnography of communication: “[t]hereby, the hard stuff of phonology, morphology, syntax and semantics is avoided; worse, the hard stuff of pragmatics, ethnography, statistics, data collection, and the like is often also not in evidence” (2009: 99). Moreover, he isolates one important misunderstanding in the conversation between SLA and sociolinguistics: namely, the former field reduces the concerns of the latter to what might be called “socially sensitive pragmatics” (155).

A passing glance over the literature shows this version of the misunderstanding to be at least somewhat well-represented. For example, Oxford (2005: 245-252) describes the goal of her chapter (in an applied linguistics handbook), titled “Sources of Variation in Language Learning,” as an essay which will summarize the research on sources of variation in language learning. Oxford’s categories include (a) ‘large culture’ (e.g. individualist vs. collectivist cultures); (b) ‘small culture’ (autocratic vs. democratic/participatory teaching approaches); (c) second or foreign language learning

---

5 Preston (1989), however, explicitly notes that to draw analytic brush strokes in terms of individuals vs. interactions is arbitrary; indeed, “[a]s interrelated variables multiply, it will become as difficult to refer back to some as it would have been to exclude them at an earlier stage of consideration. Perhaps only the fact that interactions are made up of individuals suggests this order” (53)
environment (English learning in the United Kingdom vs. English learning in Egypt); (d) stylistic factors (learning styles, personality types); (e) cognitive and affective factors (motivation, anxiety) and (f) demographic factors (gender, age).

Young (1999) notes that approaches like these—including ethnomethodology and conversational analysis—which use pre-existing social categories to characterize emergent and dynamic contexts, are widespread in research on sociolinguistics and SLA. Here, fixed factors including gender, age, proficiency are “relevant in understanding a focal instance of language use only to the extent to which participants orient themselves to those categories in interaction” (1999: 106-107). Young characterizes this ‘hermeneutic approach’ to SLA (Markee 1994) as the framework which has “emerged as the dominant tradition in sociolinguistic research in SLA over the past five years” (1999: 107).

More variationist-oriented approaches SLA have provided different ways of generalizing the nature of interlanguage variability by proposing factors which tend to be linked either to the linguistic system itself or to the speakers’ social situation. This taxonomy has produced the use of terms like VERTICAL vs. HORIZONTAL variation (Corder 1977, 1981), ‘individual characteristics’ vs. ‘interactional factors’ (Preston 1989) or ‘low-level’ vs. ‘discourse-level’ constraints (Adamson et al. 1996). Accordingly, two primary volumes on research on variation in second language published in 1987 (eds. Gass, Madden, Preston and Selinker) have the (symmetrical) subtitles Psycholinguistic Issues and Discourse and Pragmatics.
In the years since the publication of those two volumes, their distinctions have increasingly seemed less useful for the task of providing unified accounts in a workable descriptive model of a of second language variation. Tarone (1989) and Young (1996), for example, show the importance of working across both ‘high’ and ‘low-level’ sets of constraints. Tarone (1989) investigates task-based variation of four grammatical variables: 3rd person singular -s, the article, the noun plural -s, and 3rd singular direct object pronouns in L1 Japanese and Arabic interlanguage. Different grammatical forms seemed to show conflicting behavior across tasks which ostensibly required less- vs. more- attention to form: in the same formal task (a grammar test), noun plural -s did not shift; third singular -s seemed to increase in accuracy, while articles decreased in accuracy. A reanalysis of her original coding scheme, taking into account two new factors, improves the predictiviness of the analysis:

(1) the connectedness of the discourse required by the task, and
(2) the communicative pressure brought to bear upon the speaker to be clear in transmitting information (pp. 11-12)

Tarone notes, for example, that the grammar test, which consists of isolated sentences and thus requires no cohesiveness at all, constrasts in terms of connectedness and communicative pressure (i.e. to be clear) with the narrative task, which requires consistent cohesiveness and a listener with an immediate need for information. The grammatical structures under investigation facilitated this clear communication to various degrees: 3rd singular -s, for instance, is usually redundant in terms of coding
subject information, while articles (e.g. the choice of *a/an* vs. *the*) provide a rich tapestry
of discourse-level meanings related to what is known to the speaker, addressee, or both.

In this sense, the Japanese and Arabic speakers formed a single speech
community oriented to the same norms: connected discourse when it is required and
clear communication when there is a listener involved. In other ways, perhaps linked to
transfer, learners from such typologically divergent language groups may constitute
distinct, L1-specific speech communities. In this way, Tarone locates patterns of
variation located at the individual- vs. group-level, considered in terms of what kind of
speech communities these individuals form within and across language backgrounds
(Bayley and Preston 2009).

Bayley and Preston (2009) view interlanguage through the lens of these distinct
types of learner speech communities, an approach which may be more more workable
than a *post-hoc* description of low- vs. high-level factors governing interlanguage
variation. This type of reframing may be a familiar reflex, however: processes which
were once considered ‘external’ and ‘internal’ (competence vs. performance) were
unified by descriptions of a speech community, constituted by intersubjective
competence, as in New Yorkers’ variable productions of */r/ (Labov 1972b). With the
establishment of these empirical facts, there was no longer such a clean line between the
internal and the contextual, between the ‘psycholinguistic’ and the ‘attitudinal,’ if
productions depended not only on linguistic forms and organization, but social
information was linked to those forms.
As such, the role of the group and the individual (and their attendant linguistic systems) has been a proving ground for variationists in the context of its disciplinary trajectory (Hymes 1966, Hymes and Gumperz 1972, Labov 1966, Labov et al. 1968, Wolfram 1969, Gumperz 1982 with respect to more abstract, univeralist explanations for the organization of linguistic systems (Chomsky 1965 Pinker 1994), leading to the construction of an ‘ideal speaker-listener’ as the empirical source for linguistic inquiry (Chomsky 1965: 3-4). What has been convincingly established for NS speech within the variationist paradigm, the central reality of intersubjective knowledge, has not yet been digested by much of the research in SLA, whose focus remains the internal knowledge of an individual language learner. Here, the methodologies for eliciting and analyzing data of SLA and variationist researchers are often mutually exclusive, not simply incompatible. Though there have been relatively more group studies in the last few decades (Gass and Selinker 1994; Larsen-Freeman and Long 1991; Pica 1997a) SLA research has often focused on the learner as an individual—or, in some circles, on the individual learner’s ‘knowledge’ as competence (Tarone 1990). Variationist sociolinguistics, on the other hand, has its origins in large groups of aggregate data from speech communities in large urban centers (Labov et al. 1968, Wolfram 1969). Moreover, Bayley and Preston (2009) point out that a general misunderstanding of sociolinguistics in SLA circles (integrating ‘socially sensitive pragmatics’ into language learning) often

\[7 \text { It has been shown in SLA studies, of course, that individuals can have widely different routes to proficiency (Bialystok 1990; Langman 1998; ), to say nothing of their idiosyncratic strategies for voicing ‘styles and selves’ (Johnstone 1996; Bucholtz 1999b; Eckert 2000). These concerns are set aside for now.} \]
leads to studies where “the hard stuff of phonology, morphology, syntax and semantics is avoided; worse, the hard stuff of pragmatics, ethnography, statistics, data collection and the like is often also not in evidence” (p. 99).

If the role of the individual and the group (i.e. as a Labovian ‘speech community’ characterized by ordered heterogeneity in not only socially significant meanings and situations but complex linguistic variation in linked forms) is key to bridging the theoretical and methodological gap between variationist sociolinguistics and SLA, then understanding exactly what kinds of groups we consider—their composition, the factors that link their members, etc.—is crucial. Preston and Bayley (2009: 101) suggest that a central task of variationists working in SLA is to provide an empirical base for three conditions:

1. All learners from the same language background make up learner communities, and
2. All learners from all language backgrounds belong to the same learner community, and
3. Subgroups of learners even from the same language background make up distinct communities

The linguistic correlates of conditions (1), (2), and (3) would be, in terms of interlanguage variation, processes of 1. transfer 2. universals and 3. individual variation. How could these conditions be spelled out empirically? In the next section, we will move to an exploration of individual studies using variationist tools to explore multiple independent influences on variable tense-marking in second language speech. This group of studies has helped frame what categories to look for in variation in past tense marking, how to analyze these categories, and how an analysis may bridge both
multiple linguistic and social levels (i.e. how a variable form operates for learners from
different language backgrounds or Lengths of Residency).

2.2 Tense unmarking: The variationist rubric

2.2.1 Wolfram et al.’s Vietnamese English studies (1980s)

The variationist study of L2 tense marking was initiated by Christian, Wolfram, and
Hatfield (1983), and reported in Wolfram and Hatfield (1984), Wolfram (1985) and
Hatfield (1986) in a project which has come to be known as the Vietnamese English (VE)
studies. These studies investigated the systematic constraints on tense marking in a
Vietnamese community in Northern Virginia. The authors used data from 90
sociolinguistic interviews with subjects from four age groups (10-12, 15-18, 20-25, and
35-55) and two length of residency (LOR) groups (a subsection of 32 speakers from the
original sample was ultimately used for some analyses). The VE studies explored the
influence of eight factors on the shape of morphological tense marking in English:

1. Regular forms:
   a. /t/ following voiceless stops which are not alveolar, as in /mIst/ ‘missed’
      or /kIkt/ ‘kicked’
   b. /d/ following a voiced stops which are not alveolar, as in /peyd/ ‘paid’ or
      /lind/ ‘leaned’
   c. /ld/ following alveolar stops, as in /tritld/ ‘treated’ or /reydld/ ‘raided’

2. Irregular forms
   a. suppletive forms (e.g. go/went, am/was)
   b. internal vowel change plus a suffix (e.g. keep/kept, tell/told/brown)
   c. ; also modals like can/could and will/would
   d. internal vowel change (e.g. come/came, run/ran)
   e. modals (a special case of internal vowel change and suffix addition:
      can/could, will/would)
   f. replacive final consonants (have/had, make/made)
The results demonstrated that both the phonetic composition and the phonological environment of the past tense form systematically affected the relative frequency of marking across both LOR groups. First, regular forms are less likely to be marked than irregular forms, a pattern found in both second language (Dulay and Burt 1974, Ellis 1987) and first language (Brown 1973) acquisition.

Second, surface umarking is more likely for forms ending in a phonetic cluster (/mIst→ mIs/ ‘missed’ or /kIk→ kIk/ ‘kicked’) as opposed those ending in a singleton consonant (e.g. /peyd→ pey/ ‘paid’). Finally, cluster forms preceding a word which begins with a consonant (e.g. missed school, kicked people) are more likely to be unmarked versus those forms which precede a vowel (e.g. missed autumn, kicked air). This systematic process of cluster reduction has been widely documented across dialects of English (Wolfram and Fasold 1974, Guy 1980), including Spanish-influenced varieties of Tejano English (Bayley 1994), Los Angeles Chicano English (Santa Ana 1996, Fought 2003), and Puerto Rican English (Wolfram 1974).

Furthermore, the shape of the irregular past tense forms also seemed to constrain marking patterns in a systematic way; this effect also held for both LOR groups. Here, suppletive forms (e.g. go/went) are most likely to be marked for past tense and replacive forms (e.g. have/had, make/made) are least likely to be marked. Overall, irregular forms (i.e. as a class) were marked more often than regular forms.

These results are summarized in Figure 1 below (from Wolfram and Hatfield 1986: 21).
Figure 1: Incidence of unmarked tense (Unm) for types of irregular verbs, by length of residency (LOR). (Reprinted from Wolfram and Hatfield, 1986, p. 21.)

This graph displays the incidence of unmarking based on irregular verb type, where Regular Forms (1a.-c.) > Replacives (2e.) > Modals (2d.) > Internal Vowel Change (2c.) > Internal Vowel Change plus Suffix (2b.) > suppletives (2a.). This hierarchy is explained on the basis of the PRINCIPLE OF PERCEPTUAL SALIENCY, in which “the more distant phonetically the past tense irregular form is from the non-past, the more likely it will be marked for tense” (Wolfram 1985: 247).

Though the principle holds reliably at longer LORs, Wolfram notes a strong lexical constraint in the early stages of acquisition.8 There appears to be considerable

---

8 Wolfram, Carter, and Moriello (2004) also highlight the role of the frequent lexical items in the early stages of acquiring phonetic processes: in their study of emerging Hispanic communities in North Carolina, they note “some speakers may acquire a glide-reduced production of the /ai/ vowel in the lexical item Carolina well before or even while resisting the acquisition of a generalized version of prevoiced glide weakening” (p.353).
individual variation in the rote forms a beginning learner acquires: one learner may acquire marking for the irregular past tense on *go/went* while ignoring *come/came*; another learner may pay attention to marking *come/came* while ignoring *go/went*. Finally, an unrelated type of lexical effect shows up in marking patterns related to relative verb frequency: at particular points during the acquisition process (e.g. internal vowel change for 1-3 LOR group and replacives for 4-7 LOR group), frequently used verbs favor tense marking. Overall, the VE studies do not find substantial evidence for discourse-level constraints operating independent of surface level linguistic favors.

We will return in section 1.5. to a discussion of how the data from the VE studies fulfill the requirements for Preston and Bayley’s three types of speech communities (described in section 1.1.), i.e. how speakers in this study belong to a group of 1) L1 Vietnamese learners of English past-tense forms 2) all learners of English past-tense forms, as well as 3) subgroups of L1 Vietnamese learners (e.g. distinguished by proficiency) of English past-tense forms.

2.2.2 The Chinese English studies (1990s)

Bayley (1994) follows up on the VE studies in a comprehensive analysis of the past-tense marking patterns of 20 adult native speakers of Mandarin who are learning English while living in California. Overall, the Chinese-English data provide additional (crosslinguistic) evidence for the influence of the principle of saliency in L2 tense-marking patterns, while expanding the analysis to consider additional factors such as grammatical aspect. As we will see, Bayley’s studies ultimately provide clues for how to
unify our investigations of learner variation through an identification of common linguistic processes and speech communities \((1) - (3))\) constituted by one set of L2 speech data.

Speakers in this study were recorded during 2 sociolinguistic interviews: a one-on-one with the researcher, a non-Chinese native speaker of English, and then in a dyad with another informant. The effects of seven factors on over 5000 tokens were then tested using a VARBRUL analysis:

(a) verb type, or phonetic form of the past tense  
(b) the preceding segment (applies only to regular nonsyllabics and some replacives)  
(c) the following segment (applies to regular nonsyllabics and some replacives)  
(d) grammatical aspect  
(e) English proficiency (Test of English as Foreign Language (TOEFL) 550 +, TOEFL 510 -)  
(f) participation in English-speaking social networks (mixed social network, predominantly or exclusively Chinese social network)  
(g) interview type (individual, paired)

After collapsing a few of the VE coding categories for verb type,\(^9\) the results confirm the effect of saliency: that is, a greater phonetic difference between a past and present tense form promotes marking of a past-reference verb for tense. Here, only weak syllabics deviate from the saliency hierarchy; this result is explained on the basis of stress: 85\% of the regular (cluster) nonsyllabics in the corpus have –t/-d affixed to a stressed syllable, while the –\textit{ed} syllable (as in \textit{wanted}) is always unstressed.

---

\(^9\) Bayley combines 1. strong verbs and copulas other than first person singular, as well as 2. replacives and weak nonsyllabics, citing a small difference in factors values. He justifies these decisions by noting that both verb types in 1. involve a vowel change without affixation of a regular past tense marker (\textit{come} \rightarrow \textit{came}; \textit{is} \rightarrow \textit{was}) and 2. involve a change in the final segment (e.g. \textit{send} \rightarrow \textit{sent}; \textit{show} \rightarrow \textit{showed}; \textit{walk} \rightarrow \textit{walked}). The pre- and post-collapse analyses do not differ significantly with respect to goodness of fit (p > .25).
Proficiency level, however, helps sift out some interesting effects. Overall, the ordering of constraints remains relatively stable at higher and lower proficiency levels: the more salient verbs favor marking, while the less salient verbs disfavor marking. There is an increase in marking overall at the higher proficiency levels, however, these marking patterns do not proceed evenly across verb types; rather, Bayley notes, “the salient verb types lead” (1994:172) such that advanced learners mark 80% of suppletives (vs. less advanced speakers’ 43%), a difference of 37%. The corresponding difference between weak syllabics marked at higher and lower proficiency levels is only ten percent. As acquisition proceeds, variability increases across morphological classes, resulting in a greater “spread” for higher-proficiency learners who increasingly articulate morphological distinctions in their marking patterns.

In addition to phonetic composition, phonological environment of the verb form also systematically constraints marking patterns. In general, -t, d is more likely to be deleted if it is preceded by an obstruent than a liquid or a vowel (cold vs. las’); furthermore, deletion occurs on average more often if the following segment is an obstruent or liquid vs. a vowel (las’ night vs. last apple). These patterns are well-documented in native varieties of English (Guy 1980, Labov 1989) as well as in the nonnative (VE) data of Wolfram and Hatfield (1984).

Finally, Bayley finds that grammatical aspect strongly constrains marking patterns across morphological classes of verbs, with perfective aspect favoring (factor weight=.68) marking over imperfective aspect (factor weight=.32). As with saliency, this
factor seems to constrain the variation learners’ interlanguage at a steady rate throughout the acquisition process; perfective verbs favor marking and imperfective verbs discourage it at approximately the same probabilistic weight. In other words, as learners acquire more English, the rule does not disappear in the evolving interlanguage: in fact, as Bayley puts it, advancing speakers simply “turn up the input probability as they move in the direction of target language norms” (1994:175). This effect holds even when data are reduced to the individual level, such that there is a stable relationship between factor values for perfectives versus input probabilities for 15 speakers in the study.

Furthermore, Bayley observes that the stable influence of perfectivity convincingly recapitulates the behavior of the classic sociolinguistic variable (e.g. (r) in New York City) as described in Labov (1972b): here, the aspectual constraint produces a “uniformity of abstract patterns of variation which are invariant in respect to particular levels of usage” (p. 121). In other words, language learners themselves exhibit systematic linguistic variability characteristic of a NS speech community. As such, Bayley argues that the results shown here have two important implications:

First, the path of acquisition for this feature is tied to markedness. Second, and more interesting, the relatively level influence of past-tense marking according to perfectivity across proficiency levels suggests that no radical restructuring of the grammar as regards this feature has taken place. In other words, although the high-proficiency respondents in this investigation mark more pasts, the probabilistic weight assigned to one of the factors (perfectivity) that significantly influences this marking is nearly equal to the weight assigned the same factor for lower-proficiency speakers from the same group of learners (p. 100).
Thus, in addition to confirming principle of saliency established in the VE studies, Bayley posits that the influence of perfectivity, also stable across proficiencies, may be central (i.e. present across L1s) in accounting for past tense marking patterns.

2.3 Tense unmarking studies: The discourse rubric

2.3.1 Foreground vs. background

In contrast to variationist analyses, which focus on low-level linguistic factors, a strand of SLA studies has turned to higher-level factors, including discourse constraints, to account for patterns of variability in tense marking for speakers from a variety of source/target language pairs (Larsen-Freeman 1980; Kumpf 1984; Reinhart 1984; Meisel 1987; Véronique 1987; Bardovi-Harlig 1994,1995). This approach assumes that the organization of discourse, typically, but not always, investigated in the context of a narrative (Labov and Waletzky 1967, Schiffren 1981) is the primary motivation for tense alternation in interlanguage speech:

The assumption is that any grammatical form appears to fulfill a function in the discourse: it is the discourse context which creates the conditions under which the forms appear, and in order to explain the forms, it is necessary to refer to this context (Kumpf 1984: 132).

Specifically, these analyses take as primary the division of ‘foreground’ and ‘background’ that had been described in functional accounts of grammar (Hopper 1979, Givón 1982). In general terms, background clauses set the scene, change the normal sequence of events, give descriptions, and make evaluations. By contrast, the foreground is the primary event line that advances the storyline as it unfolds: it tells ‘what happens
Adamson (2009) demonstrates how foreground and background clauses might be identified in a narrative in Figure 2:

Figure 2: Sample of narrative text from Adamson (2009: 64) coded for foreground vs. background

Beginning with Kumpf’s (1984) study, which examined the speech of Tomiko, a native Japanese-speaking adult, a strand of research in the 1980s examined the extent to which past-tense marking in interlanguage may describe a foreground/background axis rather than a temporal one (Tomiko unmarked tense in all foreground clauses). In a study of native Spanish-speaking students at a middle school in Arizona, Adamson et al. (1996) found evidence that speakers unmarked past tense more frequently in foreground clauses (especially at lower proficiency levels), but noted they did not do so at the dramatic levels of Tomiko.


2.3.2 Function of the historical present in narratives

On several levels, which will be explored in this section and the next, the foreground/background account is not as straightforward as it may appear to be. Several authors have pointed out wrinkles in the new discourse-level overlay.

First, Wolfson (1982) argues that NS intuitions for temporal reference may not capture the rules that non-native speakers use in narrative discourse. She elaborates on the use of the conversational historical present (CHP) as a sociolinguistic (interactional/discourse-level) variable which alternates depending on constraints in the speech situation, as well as a textual device for organizing the (native) speaker’s narrative, specifically, the separate episodes in the storyline from one another. Following Godfrey (1980), and working within the narrative framework provided by Labov and Waletzky (1967), Wolfson argues for

…the pattern past-CHP-past with the partition of events being defined by the switch in verb tense such that the most dramatic point is signaled by a switch to the past tense (1982: 63)

Wolfson illustrates this function with the following excerpt from her data, a “fight story” recounted by a teenaged speaker of African American English. The italics are in the original excerpt; CHP is underlined here for convenience:

M: “You know why they jumped him?”
P: “Well, you see there was a Lincoln boy and a Snyder boy got in a fight. And Benny was about my height. And Benny ran and he grabbed the two and he pulled them apart. And, like he got them both. And like the Snyder boy started to hit Benny and Benny pushed him back and he fell on the ground. And the dude from
Lincoln just *stand* there and *looked* at him because he *know* how Benny is. So um, Benny *lets* him go and Benny *turns* around and *says*, ‘Why you wanna do that?’ And when Benny *said* that, all the other dudes just *jumped*. Well, when they *jumped*, they just, all the other kids *torn* them off Benny. Like see, Benny *recognized* most of the kids that *was* on him, *hitting* him. And like Benny just *went* out and *got* his brothers and that *was* it that night (1982: 62).

Here, CHP serves the function of bringing the dramatic climax of the narrative (“Benny lets him go and Benny turns around and says…”) to the forefront of the interaction in terms of a temporal shift in morphology.

In order to consider the effects of these types of episode boundaries, Wolfram (1984) re-tabulates representative data from one speaker in his VE study. He finds a few cases which support switches at episode boundaries, but concludes on the basis of more prominent low-level patterns (e.g. consistent unmarking of *have* and *go*; consistent marking of *come* and *do/don’t*) that a consideration of the surface constraints is critical when examining tense alternations.

### 2.4 Tense unmarking studies: The role of (two kinds of) aspect

The acquisition of tense and aspect has been an active, contentious area of research in both first-language acquisition as well as SLA. In this section, a discussion of that debate will bring the issues into contact with findings on interlanguage tense/aspect in the variationist literature. More specifically, we will consider which aspectual meanings overlay the constraints proposed so far in variationist studies (i.e. are

---

10 Wolfson codes cases like these as ambiguous, as it is unclear whether they are unmarked for third singular or past.
ultimately isomorphic), and, by contrast which distinct meanings/constraints may emerge (or be usefully recast) by looking across disciplines.

2.4.1 Semantic/lexical constraints: Akionsart

Vendler (1967) classified verbs on the basis of lexical aspect, or akionsart, an inherent property of their meaning, proposing a four-way taxonomy. This type of aspect has also been called “situational aspect” (Smith 1983). Verbs (and sometimes their predicates) fall out into distinct categories according to aspect: STATES, ACTIVITIES, ACCOMPLISHMENTS, AND ACHIEVEMENTS. These categories may be further sorted by which semantic features attach to each classification: stative verbs (like, think, have), for example, have no endpoint: they are STATIC (vs. DYNAMIC). Activity (or unbounded) verbs (walk, drive, swim) have duration; however, they are temporally homogenous: they have no natural endpoint in time which logically defines them. Accomplishment (or bounded) verbs (make (a speech), build (a house), run (a mile)), however, have a necessary endpoint: they are TELIC, as opposed to activities which are ATELTIC. Finally, achievement verbs (arrive, find, die) differ from all three previous types in that they are PUNCTUAL or instantaneous: at the moment they occur, the action to which they refer is finished. As such, they have a natural endpoint: like accomplishments, they are necessarily telic.

Vendler provides several tests for each kind of classification: the basic meanings of activity and accomplishment verbs, for example, may occur in imperatives (e.g. Run!,...
Run a mile to the store!) while stative and achievement verb meanings cannot (*Know the answer! *Happen!)

In both the L1 and L2 acquisition, it is consistently observed that speakers sometimes mark aspectual designations over temporal reference in the early stages of development (Antinucci and Miller 1976, Bloom, Lifter, and Hafitz 1980, Weist et al. 1984, Slobin 1985, Andersen 1989, 1991; Shirai and Andersen 1995, see Andersen and Shirai 1996; see Bardovi-Harlig 1995a for a review). De Villiers and de Villiers (1985), for example, found that in the early stages, children acquiring English used the past tense to mark perfective and punctual actions; similar findings have been produced for French (Bronckart and Sinclair 1973) and Modern Greek (Stephany 1981). SLA research has tested this claim, termed the PRIMACY OF ASPECT HYPOTHESIS (POA)\textsuperscript{11} across a variety of L1/L2 typologies, including Russian- (Flashner 1989) and Spanish-speaking (Robison 1990, 1995) learners of English, as well as Chinese learners of Japanese (Shirai and Kurono 1998). Accordingly, Bardovi-Harlig and Bergström (1996) argue that the developmental sequence, in which past-tense marking begins with achievement and accomplishment verbs and progressive starts with activity verbs, is becoming established as a universal in SLA research.

2.4.2 Grammatical constraints: Point-of-View aspect

As opposed to lexical aspect, which is tied inherently to the verb’s meaning, grammatical aspect describes a particular speaker’s point-of-view on a specific event

\textsuperscript{11} The stronger version, termed the ‘Defective Tense Hypothesis’ emerged in Weist et al. (1984)
described in a sentence (or text). Comrie (1976) describes this function as different ways of “viewing the internal temporal consistency of a situation” (p. 3): in English, for example, a speaker may view an event in its entirety as a single whole (perfective aspect: *She had lunch*) or, alternately, may attend to the internal structure of an event (imperfective aspect: *She was having lunch*). The imperfect-preterite distinction is morphologized in many languages, such as French, Spanish, Russian, Modern Greek and Persian.

Comrie’s (1976) binary features constituting to grammatical aspect are mapped on to Vendler’s four-way classification schema for lexical aspect in Table 1:

**Table 1: Semantic Features for Vendler’s (1967) four event classes**

<table>
<thead>
<tr>
<th></th>
<th>State</th>
<th>Activity</th>
<th>Accomplishment</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctual</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Telic</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dynamic</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Know</td>
<td></td>
<td>Run</td>
<td>run (a mile)</td>
<td>Arrive</td>
</tr>
</tbody>
</table>

As mentioned in section 1.2.2., Bayley (1994) tested the perfective-imperfective aspectual opposition for L1 Mandarin learners of English. He noted that Chinese grammaticalizes aspectual distinctions over tense, marking the perfective with the clitic particle *-le* (Li and Thompson 1981). Bayley concludes that marking patterns in Chinese-English interlanguage, where perfective aspect favors marking, are the result of the convergence of two processes: 1. a typological universal (single, completed events will prototypically occur in the past) and 2. transfer from the first-language aspectual system (which marks perfective with *-le*).
Bayley’s analysis is in line with Andersen and Shirai’s (1994, 1995) influential prototype account, which addresses the POA Hypothesis by proposing that “children acquire a linguistic category starting with the prototype of the category, and later expand its application to less prototypical cases” (p. 758). Andersen and Shirai try to account for the (sometimes conflicting) observations in the SLA literature related to lexical and grammatical aspect, as well as foregrounding and backgrounding functions in interlangauge speech, by relating aspectual meanings to prototypical characteristics of ‘past-ness’ (including, for example, telicity). Figure 3 provides an example of how prototypes are constituted using basic semantic feature related to tense and aspect.

![Prototype Diagram]

**Figure 3: The relationship between perfect and past (reprinted from Andersen and Shirai 1994: 149)**

The speakers in the first prototype studies (reanalyses of the speech of Brown (1973)’s Adam, Eve, and Naomi) initially mark with past morphology with instances of [+punctual], [+telic], and [+result] verbs, and gradually extend past inflections to other peripheral/non-prototypical verbs. Furthermore, a distributional bias (in the direction of
prototype features, though not reaching the levels of the children’s marking patterns) is noted in the input, the native speech of the children’s mothers (Table 2).

**Table 2: Inherent Aspect with Past and Progressive Inflections in mother’s speech (reprinted from Shirai and Andersen 1995: 751)**

<table>
<thead>
<tr>
<th>Mother</th>
<th>State</th>
<th>Activity</th>
<th>Accomplishment</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam’s mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>past</td>
<td>17%</td>
<td>8%</td>
<td>11%</td>
<td>64%</td>
</tr>
<tr>
<td>-ing</td>
<td>0%</td>
<td>56%</td>
<td>11%</td>
<td>33%</td>
</tr>
<tr>
<td>Eve’s mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>past</td>
<td>13%</td>
<td>7%</td>
<td>21%</td>
<td>59%</td>
</tr>
<tr>
<td>-ing</td>
<td>0%</td>
<td>53%</td>
<td>14%</td>
<td>33%</td>
</tr>
<tr>
<td>Naomi’s mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>past</td>
<td>12%</td>
<td>18%</td>
<td>12%</td>
<td>58%</td>
</tr>
<tr>
<td>-ing</td>
<td>4%</td>
<td>61%</td>
<td>12%</td>
<td>23%</td>
</tr>
</tbody>
</table>

The prototype analysis is attractive in unifying discourse-level (narrative) functions such as foregrounding and backgrounding with the optimal kinds of semantic and grammatical ‘tools’ speakers (both native- and non-native) have to code these functions. In addition to the explanation for foreground and backgrounding, it brings together many of the important observations noted in the previous sections, including 1. the stable influence of perfectivity (cf. Callahan 2008) and 2. Andersen and Shirai (1994) comment on the role of frequency on past-tense marking patterns:

It is almost common sensical to observe that learners cannot learn everything simultaneously and instantly. In the realm of tense and aspect marking, it appears that what they first learn are the basic tools they will continue to use later as more fluent speakers. We content that learners are motivated by the same communicative need to distinguish central events (with past/perfective marking) from simultaneous situations (with progressive) and from both static background (initially with present verb forms, much later with past imperfectives, or, for languages like English, with past habitual would/used to). Furthermore, as learners acquire more of the linguistic repertoire they need to use the language they are acquiring, they elaborate on this basic framework rather than abandoning it (p. 153).
This analysis becomes more attractive in accounting for several other recent findings that are relevant to variationist work on interlanguage. First, the prototype account has been applied to the tense-aspect systems of creoles, which tend to have an overt imperfective or progressive marker, and encode perfective or nonprogressive with zero (Bickerton 1981, Shirai and Andersen 1996, Poplack and Tagliamonte 2001). Andersen and Shirai (1995: 760) also note that in language change, past tense and perfective morphology have been documented to have developed from [+resultative], [+perfect] aspect markers in languages of the world (Bybee and Dahl 1989, Bybee et al. 1994).

Finally, Bayley (1999) and Schecter and Bayley (2002) extend the POA hypothesis to communities undergoing language shift, as from Spanish (/English bilingualism) to English in south Texas. In a study of 27 elicited narratives from children aged 4-12, Bayley investigates the extent to which loss of tense/aspect forms mirror the language acquisition of L2 (Anglo) learners of Spanish. He finds that the prototype account predicts how morphological distinctions may be lost during the process of language shift (Table 3). Specifically, there is an implicational relationship for the order in which non-prototypical aspectual forms loose their past-tense marking.
Table 3: Preterit tense by Aspectual Class in Mexican-Origin Children’s Spanish Narratives (adapted from Schecter and Bayley 2002: 125)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Punctual</th>
<th>Telic</th>
<th>Activity</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>+</td>
<td>(-)</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>(-)</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td>+</td>
<td>+</td>
<td>(-)</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>+</td>
<td>(-)</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>+</td>
<td>(-)</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

2.5 Next steps

In looking toward the future, Wolfram (1978: 19) argued that similar systems of structured, inherent variability exist in both ‘sociolects’ of NS speech and in the interlanguage systems of language learners. Within both systems, he hypothesized:

…there is a patterned relationship in which certain variants are clearly favored over other variants depending on the context. And although there may be
important nonlinguistic variables that influence relationships of more and less (e.g. stage of second language acquisition), there also is evidence for the existence of independent linguistic constraints on variability (e.g. environment).

Wolfram’s hypotheses have borne fruit in our evolving understanding of tense/aspect interlanguage variation. Specific developments include:

1. Unified accounts of tense and aspect marking for native and NS speakers, specifically, those accounts (including POA and prototype models) which also capture processes in discourse structure (foregrounding/backgrounding), creolization, language change, and language shift

2. The role of converging linguistic factors at the phonetic, morphological, and phonological levels of L1 and L2 speech, including the PRINCIPLE OF PERCEPTUAL SALIENCY (see section 2.2.1)

3. The nature of interlanguage variation located at the individual- and group-level, especially in terms of what kinds of speech communities these individuals form within and across language backgrounds

In the thirty years since Wolfram’s predictions, in fits and spurts, SLA and (variationist) sociolinguistics have begun to glance across the Saussurian divide: to begin to provide converging accounts of social and psycholinguistic processes in both first- and second-language acquisition (Preston 1989, 2009). In important ways, this convergence has occurred to the extent that 1. variationists were able to join larger disciplinary conversations on a cognitive capacity for language-making and 2. SLA
researchers were able to compromise on their categorical intuitions and peremptory conditions in favor of empirical facts for describing the language usage of real people living in communities in time and space.

Along these lines, Wolfram concludes:

I am convinced that many of the questions about linguistics in general and contrastive linguistics in particular will not be answered until we look at language in terms of its actual usage rather than some idealized construct of how we expect it to work prima facie” (1994:197) [my emphasis].

It is hoped that the present study, an account of interlanguage past-tense marking in the emerging Hispanic English of one community in North Carolina, will continue the productive, cross-disciplinary conversation between variationists and researchers in SLA.
3. Traditions in variationist sociolinguistics

In this section, I will provide an overview of some of the assumptions and methodologies used in quantitative variationist (socio)linguistics, beginning with classical variable rules and ending with contemporary questions of how to best hypothesize about, code, and analyze grammatical variation in the ethnolect. A familiarity with these principles will allow the reader to understand the rationales for general design of this study, including coding decisions (sections 4.4.1-4.4.5), presentation of results (section 5.), relevance of the results in the triethnic community context (section 6) and broader applications for the general findings (section 7).

Ultimately, along with the presentation of results in section 5. and 6., I hope to establish some of the ways in which variationist methodological practice establishes and reifies assumptions in terms of core disciplinary constructs (e.g. formal variable rules, VARBRUL ‘factor groups’), and how a shift to more recent techniques and models (comprehensive statistical programs, usage-based approaches) can determine how methodology itself dictates the kind of results we find.

3.1 The sociolinguistic variable

Traditionally, variationists have studied how language structures vary systematically at various levels of linguistic organization (for example, the syntactic, phonological, lexical level, or). The task of a variationist typically involves isolating a language structure, termed a linguistic variable, identifying variants of the variables, and investigating how the variants of the structure patterns across different social
categories (region, age, class) and in different linguistic environments (a verb which occurs habitually vs. non-habitually):

1. The term ‘pail’ vs. ‘bucket’ for Northern vs. Southern regions of the U.S. (Kurath 1949)

2. The pronunciation singing [ŋ] vs. singin’ [n] for ‘model’ vs. ‘typical’ school-aged boy in New England (Fischer 1958)

3. The use of invariant ‘be’ in sentences like He be tired (vs. He tired) in working class vs. middle class African American Detroit (Wolfram 1969)

The examples in 1.-3. above illustrate what laypeople tend to think of as ‘dialect’ or ‘vernacular’ forms of English, or what linguists describe as examples of language variation\(^1\).

Along these lines, sections 0-3.3 will provide case studies of classic linguistic variables which have been widely studied in variationist sociolinguistics, namely copula deletion and consonant cluster reduction (CCR), in order to illustrate this complexity of dialect patterning as well as demonstrate the kinds of nuanced social effects these patterns have established in the first few decades of variationist research. After laying this groundwork, section 3.4.3 will extend these earlier methods and assumptions in order to discuss contemporary methods and paradigms for quantitative sociolinguistics, \(^1\)It is now widely accepted that all varieties of human language are equal (rule governed, systematic, and consistent) in linguistic terms. The phenomenon whereby the linguistic varieties of socially subordinate groups are stigmatized as ‘dialects’ whereas more powerful varieties are elevated to the status of ‘languages’ is identified by Lippi-Green (1997) as the Principle of Linguistic Subordination. For background on the language-dialect issue, especially in the context of African American English in the U.S., see Linguistic Society of America (LSA)’s 1997 Resolution on the Ebonics issue, as well as Shuy (1965), Wolfram (1969), Fasold and Shuy (1970).
including updated understandings of how variationists define the linguistic context (e.g. variable rules vs. usage-based effects). Section 3.5 will discuss how the unit of social organization has evolved toward the use of more dynamic categories and concepts (e.g. demographic categories vs. local social networks). Finally, these disciplinary assumptions and methodological traditions will be linked to the research questions addressed by the Durham HE data in the presentation of results in sections 5. and 6.—specifically, the degree to which Durham speech resonates with contemporary accounts of the ethnolect.

### 3.2 Case study 1: Copula deletion in early AAE

In section 3.1 (3.), we mentioned one example of a dialect structure from Wolfram’s (1969) study of African American English (AAE) in Detroit, in which a conjugated form of ‘to be’ may alternate in the sentences *He’s tired* vs. *He tired*. For the variable mentioned here (absence of the copula, or verb *to be*) numerous studies have identified a now familiar set of constraints, including following grammatical environment (*He’s [tired]adjective* vs. *He’s [a fireman]noun*), 2. following phonological environment (*He’s [t]ired vs. He’s [ə]fraid*), and 3. pronoun person and number (*He’s (3rd sg.) tired* vs. *We’re (1st pl.) tired*) (Labov 1969, Wolfram 1969, Rickford et. al. 1991).

In examples given above, from study of 2-6 year old children developing African American English in the Piedmont region of North Carolina (Van Hofwegan and Wolfram 2011), the incidence of present tense forms of to ‘be’ occurs at the highest rates with 1. a following nominal complement 2. a following vowel vs. consonant and 3. for
‘are’ (1st-3rd plural) vs. ‘is’ (3rd singular) forms. Other structural elements, including WH-question words like ‘Who’ or ‘What’ or alternate verbal complements like V+ing, as in ‘He playin’ fall in line according to the results shown in Figure 4 (from Callahan-Price 2012).

Figure 4: Percentage of copula deletion by age and structural environment

Figure 4 demonstrates that this pattern of internal variation in copula deletion shows the same rank orderings (e.g. C > V for preceding phonological environment) across different age groups (48 months vs. 1st grade) and generations (children vs. their mothers) once children have acquired the dialect constraints around 48 months.

Copula deletion in this community also illustrates important social and sociopsychological effects. For example, when children enter school, their deletion
patterns fall to levels lower than that of their mothers (while generally maintaining the same rank order), a pattern which has been accounted for by the role of school socialization in Standardized English (Craig and Washington 2006). After this initial period of low vernacularity in grades 1 and 4, the children in the study tended to follow one of two patterns, either a ‘roller-coaster’ trajectory, where overall vernacularity (in terms of dialect features per utterance) peaks at 6th-8th grade, then drops at Grade 10, or a ‘curvilinear’ trajectory, where overall vernacularity higher in earlier childhood, drop during Grade 1-4, then continues in a steady rise until Grade 10 (Van Hofwegan and Wolfram 2010) (Figure 5).
Figure 5: Alternative trajectories of change over the early lifespan: (a) ‘Roller coaster’ trajectory; and (b) Curvilinear trajectory (from Van Hofwegen and Wolfram 2011: 439)

Taken together, Figure 4 and Figure 5 illustrate how patterns of variation (in terms of copula as well as other sociolinguistic variables) can highlight both similarities and differences among and across social categories. For example, the members of this speech community show uniformity as well as distinctiveness according to the functioning of copula deletion (in the context of other sociolinguistic variables) in terms of 1.-3. below:
1. **Uniformity in overall frequencies across constraints (i.e. at one time point):**

   Children in this community tend to speak a more ‘Standard’ variety of English by first grade relative to their levels both before school entry and the levels of their mothers.

2. **Distinctiveness in trajectories of vernacularity (i.e. over many time points):**

   Some children become more vernacular as younger adolescents, then return to less vernacular levels as they age, while others become steadily more vernacular throughout adolescence with no ‘rise and dip.’

3. **Uniformity in general AAE constraint ordering:** Despite overall differences in the overall rates of copula deletion (in tandem with rates of occurrence for other AAE variables), the rank orders of AAE constraints for copula deletion remains the same across generations (mothers vs. children in the study group) and even among different generations across regions of the U.S.: for example, the same rank orders were established for preschool children in southwestern Louisiana almost four decades later (Green 2011).

### 3.3 Case Study 2: CCR across English varieties

In addition to copula deletion, a diagnostic case of variability which has been studied extensively over the last four decades is the phenomenon of consonant cluster reduction (CCR). This process occurs when syllable-final consonant clusters, usually both unvoiced ([st] as in ‘test’) or both voiced [nd] as in ‘friend’), the second of which is a stop consonant, become reduced to only the initial segment (*tes’, frien’*). This variable is
especially useful to consider in the context of the present study since its investigation serves a dual purpose. First, as a “paradigm case of systematic variability in variation analysis” (Wolfram, Childs, and Torbert 2000), the functioning of CCR illuminates much about the principles and methods of the field of variationist sociolinguistics on the whole. Second, as a phonological process, CCR is directly relevant to the grammatical process of tense unmarking since one class of regular English past tense forms ([pIkt] ‘picked,’ [lind], ‘leaned’) are formed in a final consonant clusters. Thus, the unmarking variable (‘pick’ → [pIk]; [lind], ‘leaned’) for verbal tokens in the Durham data may be due to phonological reduction in conjunction with, or in the absence of other structural factors (verb class: regular vs. irregular verb), usage-based effects (frequency of the word in discourse), or developmental influence (proficiency in English/LOR).

In fact, as a language transfer feature, CCR has had an important role in helping researchers understand the development of various ethnic dialects of English whose heritage/ancestral languages lack syllable-final consonant clusters, including African American Englishes (West African substrate), American Indian Englishes (indigenous American substrate) and Latino/Chicano Englishes in the southwest U.S. (Spanish substrate). While highlighting its contact history, CCR can also point the way towards an ethnic dialect’s future. As an ethnolinguistic marker, CCR can vary according to social factors like ethnicity, social class, gender, social networks and peer groups.

While it has been established that even formal registers of very standardized English frequently delete consonant clusters pre-consonantly (‘pick through’ for
‘picked’ through), levels of CCR across varieties dip precipitously if the cluster occurs before a vowel (‘pick up’ for ‘picked up’) or, as in the unmarking of verbs, as part of the morpheme that forms the regular English past tense marker ‘-ed’ (‘picked through/up’).

The reliable effects of phonological environment (prevocalic vs. preconsonantal) as well morphemic status (whether the cluster encodes a distinct meaning, as in the bimorphemic token ‘leaned’/’lean’ + ‘ed’ but not the monomorpheme ‘sand’) across register, social class, region, ethnicity have established CCR as a hallmark sociolinguistic variable which can reliably demonstrate nuanced social effects (Labov et al. 1968, Wolfram 1972, Fasold 1972, Wolfram and Christian 1976, Neu 1980, Guy 1991, Santa Ana 1992).

In addition to broad demographic categories like ethnicity, region, and class, CCR was shown early on to demonstrate finer-grained local distinctions. For example, Labov (1972b), in a study of the speech of adolescent male street gangs in Harlem, New York, illustrates how CCR can help delineate social networks in terms of gang affiliation. Here, not only the overall frequency of CCR rule application across constraints (phonological environment, morphemic status) but the rank orders of these constraints illustrated complex facts about social organization. In Labov’s study, the primary constraint on CCR among gang members was following phonological environment (consonants promote reduction) while morphemic status was secondary (monomorphemes promote reduction). For ‘lames,’ or individuals without strong gang

---

2 Kiparsky (1972) attributes the retention of bimorphemic vs. monomorphemic clusters to a difference in functional load, whereby inflectional misse/d/ is preserved over non-inflectional mis/t/.
ties, the pattern is reversed: their patterns of CCR respond primarily to morphemic status and not as strongly to following phonological environment.

In fact, for CCR a full range of linguistic/internal constraints function consistently across varieties of English. Wolfram, Childs and Torbert (2000); the following list summarizes these findings:

**Following context**
- preobstruent > presonorant > prevocalic
  (e.g., [bEs kId] ‘best kid’ > [bEs nem] ‘best name’ > [bEs at] ‘best at’)

**Preceding context**
- nasal > lateral > sibilant > stop
  (e.g., [wIn] ‘wind’ > [wall] ‘wild’ > [wEs] ‘west’ > [ak] ‘act’)

**Morphological marking**
- monomorphemic > redundant bimorphemic > bimorphemic
  (e.g., [gEs] ‘guest’ > [slEp] ‘slept’ > [gEs] ‘guessed’)

**Stress**
- [-stress] > [+stress]
  (e.g., [ˈkantræk] ‘cóntract’ > [kanˈtraːk] ‘contráct’)

**Social factors**
- lower social status > higher social status
- casual style > formal style
- AAVE > Anglo vernacular varieties
- Hispanicized Vernacular English > Anglo vernacular varieties
- Vietnamese English > Anglo vernacular varieties

I have examined how CCR may occur at higher rates overall before a consonant as well as when it does not encode a stand-alone meaning, as in bimorphemic tokens in which the cluster marks past tense (e.g. ‘mist’ > ‘missed’). The list shows how these constraints pattern on an even finer level, introducing the intermediate level of ‘redundant bimorphemic’ environments like sle[pt], where the cluster carries redundant...
information about past tense marking which is also signaled by a vowel change ([slip] ‘sleep’ \(\rightarrow\) [slep] ‘slept’). Even further, we can observe finer-grained phonetic distinctions: beyond sensitivity to the broad categories ‘vowel’ and ‘consonant,’ CCR patterns predictably for subcategories like a preceding nasal [wIn] ‘win’ vs. stop [æk].

Another group of internal constraints appears at the linguistic level of syllable stress, where unstressed syllables, as in [ˈkantræk] ‘contráct,’ promote reduction over stressed syllables, as in [kanˈtræk] ‘contráct.’

### 3.4 CCR as a roadmap for evolving variationist methodology

#### 3.4.1 Classical variable rules

Beginning in the 1970s, CCR was one of the first structures early variationists used in order to establish and legitimize the field as a viable subdiscipline: here, pioneering early researchers developed a conventionalized way to represent rules which specify 1. both the possible constraints a variable might operate under as well as 2. the likelihood of rule operation according to these constraints. These two processes can be formally described by 1. and 2. below:

1. \(\text{t,d} \rightarrow (\emptyset) / \emptyset / (C) \_ ## [V,C] \)

2. \(\text{t,d} \rightarrow \emptyset / \emptyset / \emptyset / (\emptyset / # / C) \_ ## [V] \)

Using a conventionalized schema, rule 1. describes how -t/-d may be optionally deleted in the following linguistic environments: 1) after a consonant (C) 2) before either a vowel or a consonant [V, C] 3) at the end of a word (##) and 4) whether as part of a
stand-alone morpheme or not (>). Rule 2., by contrast, allows us to add in probabilistic information about when the rule is more likely to be applied. As we outlined in section 3.3, clusters in monomorphemes like ‘mist’ are more likely to be deleted over bimorphemes (‘missed’); here, rule 2 uses angled brackets instead of parentheses and curly brackets and positions the environment ‘no morpheme boundary’ (∅) above ‘morpheme boundary’ (>) to indicate a higher probability of deletion for monomorphemes. The same schema is used to indicate the stronger influence of following consonant over following vowel. In this way, rule 2 is an example of the type of ‘variable rules’ that researchers developed in the first decades of quantitative variationist sociolinguistics (Labov 1969, Bailey 1971).

Variable rules were a critical first step in formalizing variable patterns as rule-governed, predictable effects vs. unpredictable reflexes of ‘performance’ (Chomsky 1965); they were also instrumental in legitimizing the speech of stigmatized ethnic groups as rule-governed, logical and consistent systems (Shuy 2003). In these ways, variable rules represented a giant leap forward. Looking back, however, it is helpful to recognize the footprint that older methodologies may have left in current modes of quantitative reasoning. For example, variable rules were typically devised in the context of an isolated, abstract linguistic ‘structure’ (e.g. CCR, copula), not a single word with a usage history and dynamic relationship to other words. A larger problem, the assumed independence of constraint variables (e.g. morphemic status, phonological environment) will be addressed by the results in section 6.1.2 (CCR in Durham HE and AAE).
3.4.2 VARBRUL

After the initial introduction of variable rules, a statistical analysis program called VARBRUL (Cedergren and Sankoff 1974) was developed to calculate probabilities, called factor weights, at a greater level of sophistication for linguistic data. Specifically, VARBRUL calculates a factor weight, or probability score (p) which predicts whether a variable rule will be applied: a value above 0.50 indicates that the factor promotes the rule and a factor weight below 0.50 indicates that the factor inhibits the rule. Another VARBRUL measure, the input probability, indicates the baseline probability that the rule will operate independently of any of the constraints under investigation (i.e. an overall tendency).

Factor weights allowed variationists not only to rank constraints (e.g. C > V) but also to compare the relative strength of factor weights among and across factors. In addition to the input probability, two other important values are associated with a VARBRUL analysis: the total chi-square and average chi-square per cell. The former value allows the researcher to assess the degree to which the factors function independently of each other: a higher score indicates the factors are significantly correlated with each other (for example, syllable stress is correlated with an upcoming pause vs. an upcoming consonant). Average chi-square per cell gives the researcher an indication of the model’s explanatory power: the lower this value (with a maximum of 1), the more indication that the model comprehensively accounts for the variation.
observed for the study variable in a given data set. Finally, as a stepwise regression procedure, VARBRUL adds in factor groups incrementally in order to assess the unique contribution of each group; factor groups which do not show significant correlations with the dependent variable are discarded from the model.

In all, VARBRUL represented a significant empirical advance over classical variable rules in that it allowed researchers to establish the significance of factors on variation as well as examine the probabilistic strength of those factors, and, to a lesser degree, the relative independence of factors (for example, in the form of total chi-square or factors which were dropped from the stepwise model). Practically speaking, factor weights and factor groups were convenient standards that could be referred to across studies, and the program itself was relatively accessible for new users. For these reasons, VARBRUL became the ‘Gold Standard’ (Guy 2010) for measuring categorical variation well into the 21st century.

### 3.4.3 Quantitative analysis Post-VARBRUL

VARBRUL has faced increasingly focused criticism in the last decade, some of it centered around the fact that it is not a suitable tool for testing interactions among factors. Furthermore, though it can be demonstrated that in some cases individual data does in fact reliably articulate group patterns, both in native speaker variation (Guy 1980) and in SLA (Bayley 1994b, Bayley and Langman 2004), the question of individual-level effects is not readily accessible when using VARBRUL to address aggregate data.

---

3Note that chi-square per cell serves a similar purpose to the $R^2$ statistic discussed in section 5.3.1.
(though each individual speaker’s data could, in theory, be entered individually). In a broader way, the debate over the quantitative sufficiency of VARBRUL spells out the evolution of contemporary questions in variationism: whereas an earlier tradition referred to static ‘features’ like CCR (or copula absence), more recent research has demonstrated the ways in which categorical variation may ultimately show continuous type effects, most vividly in terms of word frequency and patterns of actual usage.

Bybee (2000b, 2002) presents evidence that CCR is in fact not an abrupt process, but lexically gradual, as it is mediated by actual usage patterns; under this view, there is an overall ‘input value’ not only for the phenomenon as a whole, but for each individual lexical item (or combination of lexical items). This set of experiences by the speaker/listener is (cognitively) stored in the form of exemplars (Bybee 2000a, 2001; Johnson 2001, Pierrehumbert, 2001) which show association across multiple linguistic levels (e.g. phonological, morphological, and semantic similarity) and are strengthened with repeated exposure and use.

In a reanalysis of Santa Ana’s (1991) Los Angeles Chicano English CCR data, Bybee (2000b) first demonstrates that high frequency words do exhibit CCR at significantly higher rates overall (54.4% vs. 34.4% reduction; \( \chi^2 = 41.67, p<.0001 \)). Bybee (2002) then extends this account by introducing evidence that bears directly on our investigation of CCR in the context of marking a morphological distinction (regular English past tense –ed). Specifically, she shows that regular past tense forms occurred

---

4 The standard of 35 or more occurrences per million words—the median for regular past tense forms in the Francis and Kučera (1982)—is used as the cut-off point for high vs. low frequency.
before vowels at a high rate (40% vs. 21% for the overall corpus) and argues that it is the frequency of the prevocalic occurrence that conditioned retention, not morphological status as such. Thus, words that occur more frequently in a phonetic context favoring reduction undergo the variable process at a higher rate than those which do not frequently occur in the favorable context. This analysis represents a fundamental shift from the ‘functional load’ paradigm (Kiparsky 1972) in positing actual contexts of (phonetic) use in over abstract morphological information (‘bimorphemic’ vs. ‘monomorphemic’ tokens).

Bybee’s results for CCR reflect a larger debate on the role of lexical diffusion and frequency in variationist sociolinguistics—whether the word or the sound is the fundamental unit of linguistic variation and change (Cheng & Wang 1977, Phillips 1980, Labov 1989b, Labov 2010), and, consequently, how this question can be squared with traditional modes of analysis which rely on structurally-specified (VARBRUL-type) factor groups. We will address these questions empirically in the presentation of the quantitative results of unmarked tense patterns in the Durham (HE and AAE) speech community. For example, section 5.3 will outline decisions which must be made in coding and quantifying frequency in studies of grammatical variation, including whether frequency should be treated as a traditional ‘factor group’ at all, or one which intersects in complex ways with structural categories (Poplack 2001, Erker and Guy 2012). Section 6.1.2, a discussion of HE in the community context, will return to this question of usage-based effects, specifically in terms of the structural categories
morphemic status and phonological environment and their collinearity with respect to frequency.

3.5 The speech community: Intersections and innovations

In the same way that quantitative methodologies (and their underlying assumptions) have evolved over the lifetime of variationist sociolinguistics, the unit of social organization has changed from the static demographic categories discussed in section 3.3 (region, ethnicity, class, even gang affiliation) to integrate more dynamic, local and participant-driven social categories and the social meanings these participants index in discourse (Eckert 2000). This widening lens can also capture those intersections which spell out important generalities about language variation and change in social space. For example, by using a “friend of a friend” technique to access social networks, Milroy and Milroy (1978) first established how dense, multiplex ties constrained phonological variation in three urban, working class neighborhoods in Belfast, Northern Ireland. Later research (Milroy and Milroy 1993, Docherty et al. 1997) was then able to piece together a workable explanation for gendered use of more standard norms in terms of the more precise formalizations spelled out by the network dynamics of an actual community. Specifically, Milroy (1999) gives an explicit account of how male peer networks can constrain speech by preserving conservative, local norms whereas women’s broader network ties facilitate their mobility as innovators.

A subsequent strand of research using the Community of Practice (CofP) framework (Eckert and McConnell-Ginet 1992a) showed how local categories actually
constitute traditional Census’ categories like race or gender (Fought 1997) as they identities in locally-determined ways. These community-specific constructs turn out to explain variation in more useful terms than local or global characterizations alone—as in the construction of a ‘jock’ vs. ‘burnout’ (Eckert 1989) ‘nerd girl’ (Bucholtz 1999b), or Norteña/Sureña identity (Mendoza-Denton 1997).

These network and CofP studies illustrate how intersections are often the key to unlocking lower-level generalities about language variation and change. Both linguistic and social structure are coded at multiple levels of redundancy, and it has often been at the intersections of traditional and novel categories—and the conversations between them—that we find revealing patterns which drive the discipline forward to a level of greater coherence. Here, the intersections of new and old categories—network structure with gender, adolescent peer structure with class, or morphemic status and phonological environment—provide important opportunities to make higher-order generalizations about how variation and change are function in general.

3.5.1 Ethnolects

The general term ethnolect has been used to describe the English of ethnic immigrants, which may be marked by a L1 substrate influence as speakers shift from monolingualism to bilingualism (Clyne 2000). Recent research on phonological variation in multiethnic New York City (Becker & Coggshall 2009, Becker forthcoming), in northern California (Eckert 2008) and on Jewish English in the U.S. (Benor 2010) has contributed to the theme that ethnicity can be thought of as a more fluid entity than that
which has sometimes been assumed by traditional variationist accounts (e.g. of ethnic
varieties of Chicano or African American English as such).

I will ultimately make the case that Durham speakers’ shared experience with a
common set of words (and combinations of words) which they hear and use to talk
about their daily lives and shared environment (a corpus or lexicon) may outline the
most useful kind of ‘speech community’ for the results in this study. In her research on
Chicano English in Los Angeles, Fought (2006) describes this model as a “pool of
resources from which members of a speech community draw the linguistic tools they
need.” (p. 21) Here, language learners in Durham help distill some of the most frequent
(and immediately useful) components of this shared toolkit, patterns which are then not
discarded by native speakers, just supplemented and amplified. In all, the Durham data
help establish that ‘categories’ like ethnicity (Hispanic vs. African American) and
proficiency (native vs. nonnative)—and, in turn, static ethnolects—are fairly leaky
concepts, sociolinguistically speaking.

This view, which sometimes crystallizes around the construct of an
ethnolinguistic repertoire (Benor 2010), resists treating ethnic varieties as a separate
linguistic entities (Fought 2003), and instead sees variable structures as more or less
regionally and ethnically distinct resources which speakers may map on to a diverse set
of values, orientations, identities, aspirations, alignments, and day-to-day needs. For
example, in an ethnographic study of preadolescent speech in two schools in northern
California (the working-class, largely Chicano and Asian ‘Steps’ Elementary vs. the
largely Anglo, middle-class ‘Fields Elementary’ only ten miles away) Eckert (2008) shows how features like prenasal /æ/ (raised before nasals in Anglo varieties and ‘lowered’—or simply unraised—in Chicano varieties of English) become a ‘second order index’ used for purposes beyond marking ethnicity. Instead, they initiate speakers into and crowd membership and a sense of ‘coolness’ associated with a peer-based social order centered around the heterosexual marketplace.

Studies like Eckert’s which break down barriers between bounded, uniform ethnic ‘varieties’ of English are useful in many ways for considering the data in the Durham study: they help establish the playing field as a “landscape in which ethnicity plays a prominent but not determining role” (Eckert 2008: 41) and where variable features are not necessarily understood only in terms of Anglo/white English. As we will see in later chapters, features which may be thought of as distinctively African American, including habitual be and double marking of possessives (mines) occur across the board in Durham HE, by speakers of all ages, both genders, and by gang-affiliated and non-gang-affiliated students (Table 6 in section 6.1.1 shows a brief selection of identifiably-AAE features). Though more ethnographic and contextualizing work would be needed to discuss how Durham speakers construct stances, personas, styles, etc. during talk-in-interaction—as well as address the degree of automaticity involved—the data in this study shows that in terms of CCR, at least, patterns are accessible to a broad range of Durham speakers, regardless of ethnicity or even language proficiency.
In this sense, the Durham data demonstrate how a ‘pool of resources’ might be more usefully considered in terms of the local lexicon itself. As the Durham speakers acquire a particular local variety of English, e.g. that of African American peers, they must pay attention to the combinations of words around them (in other words, usage-based effects in the input). As Durham language learners keep track of how these effects intersect (for example, the intersection of phonological environment and morphemic status), then they learn speech norms in this community. In this sense, the words around them— their local, everyday linguistic experience—is absolutely critical to becoming a member of the Durham speech community as a ‘repertoire’ which is accessible to all speakers regardless of ethnicity. In fact, as we will see in section 6.1.2, both the AA and Hispanic speakers in this study chose the same combinations of words (e.g. in terms of phonological and morphemic status) to use in everyday discourse—patterns which bring to life variable effects.

The link here between local language/dialect acquisition and contemporary ‘ethnolect’ models is between language use (i.e. anyone can access sound changes in progress regardless of ethnicity) and use-age. The usage-based account (e.g. in terms of underlying interaction phonological/morphological interactions like more bimorphemes before vowels) still provides for a flexible, fluid baseline of ‘resources’ in in in terms of patterns of use. In this sense, HE speakers are not necessarily ‘accommodating’ or ‘aligning’ with their AA peers or creating unique ‘stances’ as such— they’re just sharing language/diffusing norms, and doing so regardless of ethnicity.
3.5.2 Conclusions

As a contradiction-ridden whole, the Durham speakers demonstrate a fundamental principle of the ‘speech community’ construct which remains elusive even four decades later: speakers at once share many norms, but at the same time show very distinct patterns across class membership categories. While the assertion noted in section 3.1 remains true—no two speakers, regardless of how many types of group membership they share, ever speak exactly the same way—it is simultaneously true that in order to accomplish social life through language, speakers must use and evaluate forms in recognizably similar ways:

Every concrete utterance of a speaking subject serves as a point where centrifugal as well as centripetal forces are brought to bear. The processes of centralization and decentralization, of unification and disunification, intersect in the utterance; the utterance not only answers the requirements of its own language as an individualized embodiment of a speech act, but it answers the requirements of heteroglossia as well; it is in fact an active participant in such speech diversity...Such is the fleeting language of a day, of an epoch, a social group, a genre, a school and so forth. It is possible to give a concrete and detailed analysis of any utterance, once having exposed it as a contradiction-ridden, tension-filled unity of two embattled tendencies in the life of language (Bahktin 1981: 272).

It is in this context that researchers must recognize the herculean task of second language acquisition— a process during which speakers must, as all of us do, speak both the same and differently at the same time. While no two utterances a language learner hears are ever exactly alike, he must work out which differences are special, and in what way. This requires processing multiple semiotic systems, establishing connections, uncovering regularities, then assigning those regularities reliable (and, indeed, socially
useful) meanings. Especially outside the context of formal instruction (where content meanings and rules may be specified explicitly), the language learner must accomplish a tandem task of meaning-making every time she produces an utterance: she must say *what* she wants to say while saying it the *way* she wants to say it. Even more dauntingly, these two tasks must be accomplished simultaneously—with very few do-overs.
4. Site, Speakers and Methods

In this chapter, I will discuss the site where the data for this study was collected, the general characteristics of the speakers involved, and methods used for collecting the data. I will then discuss in detail the conventions I used and decisions I made in transcribing and coding. Finally, I will provide examples of conversational excerpts with individual speakers in order to illustrate how this coding took place in practice.

4.1 Durham speech community: demographics and trends

In the past few decades, North Carolina has experienced a population boom of Spanish speaking residents which is emblematic of the southeastern United States in general. Between 2000 and 2010, more than half the increase in population in the United States as a whole was due to the growth in Hispanic or Latino origin respondents; during this period, the Hispanic population grew by 43% nationwide and by more than 3 million people, four times the growth of the total U.S. population. The group of U.S. residents who identified a Hispanic or Latino origin\(^1\) has grown fastest in the South, which experienced a 57% percent growth as a region overall from 2000-2010 (vs. 49% in the Midwest, and only 34% and 33% respectively in the West and Northeast). In fact, in the two-decade span from 1990-2000 and then from 2000-2010, growth across the South

\(^1\) The U.S. Census methodology for determining Hispanic origin has also undergone a transformation in the last half-century. Prior to the 1970 Census, the Bureau’s Hispanic ‘identifiers’ consisted of responses to questions which asked about place of birth, Spanish mother tongue, and Spanish surname. The origin self-report question was introduced in 1970, and classifies as a respondent as Hispanic or Latino if she responds that she is “a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race.” Here, the concepts of ethnicity and race are separate and distinct concepts, with Hispanic origin defined as “the heritage, nationality group, lineage, or country of birth of the person or the person’s parents or ancestors before their arrival in the United States.” Thus, residents of Hispanic origin may be of any race.
stayed in the triple digits, far outpacing traditional Hispanic strongholds including California, New York and Texas (Figure 6).

Figure 6: Hispanic Population Trends in the 1990s and early 2000s

As the bottom half of Figure 6 shows, North Carolina led the South overall in
Hispanic population growth from 1990-2010 (it had highest growth rate in the nation from 1990-2000 and the sixth-highest rate from 2000 to 2010). This growth has been constituted mostly through in-migration from speakers of Mexican (60.9% as of 2010) and Central American origin, including the three most populous Central American origin groups, made up of Salvadoran (4.7%), Honduran (3.9%), and Guatemalan (2.5%) residents.

Durham, NC is a triethnic, medium-sized city in the heavily developed Research Triangle Park region of the North Carolina Piedmont. In addition to Hispanics and European Americans, African Americans make up a sizeable portion of Durham’s population, and are in the majority of students attending the city’s public schools. Figure 7 shows the overall population of Durham (City) using self-report figures from the 2010 Census.

![Pie chart showing race distribution in Durham, NC](image)

Figure 7: Race for Durham (city), N.C. (%)

Like the city of Durham itself, the three schools chosen for this project (funded
under National Science Foundation Grants BCS-0213941 and BCS-0542139) are characterized by large African American student populations. The schools were chosen in consultation with Durham Public Schools K-12 English as a Second Language (ESL) Services staff, including Director Sashi Rayasam, a former professional contact of the author, and Principal Investigator Walt Wolfram.

4.2 Ethnographic evidence

I had personal experience working in the school system and living in the community represented by the three study sites. During 2002-2003, I taught Spanish and French (as foreign languages) at Hillside High School, a historically Black high school and social institution in Durham’s African American community (Frazier 1925). Rogers-Herr, one of schools in this study, ‘feeds into’ Hillside High according to Durham Public School’s districting plan, so many of the students I taught had attended middle school there. In addition, I was an ESL teacher at the second study site, Chewning Middle, during the summer of 2003; all of my 15 or so students that summer were L1 Spanish speakers from various parts of Latin America. During the 2003-2004 and 2004-2005 school years, I moved to a position teaching ESL in Granville County Schools, a rural county just north of the Durham city limits (one of the study sites, Chewning Middle, is located less than a mile from Granville County line). Finally, I was a resident of Durham’s Old West Durham/Ninth Street neighborhood during the period from 2001-2005. E.K. Powe, the one elementary school in the study, is located on the four-block Ninth Street business and residential corridor in Durham between Hillsborough Road
and West Club Boulevard.

The three schools have different characters, part of which is formed by their distinct demographic and geographic profiles (Figure 8).

![Figure 8: Location of 3 fieldwork sites in Durham, N.C.](image)

E.K. Powe Elementary has the highest Hispanic population (28%) and draws from five economically diverse, densely populated urban neighborhoods near downtown Durham (Watts-Hillandale, Crest Street, Walltown, Old West Durham, and the West End). Spanish-speaking parents often walk their children to school, which is located on one whole city block, and Spanish-speaking families use the school’s playing fields many afternoons for pick-up soccer games. During the time I lived and did
fieldwork near the school, its sign outside often had its weekly welcome posted in both English and Spanish.

By contrast, Chewning Middle is located in a rural setting with over 20 acres of buildings and grounds. The school is in a much less densely populated area in more remote northeast Durham, accessed most easily via Interstate I-85, and is located about a mile from both the Eno River and Falls Lake, which extends into adjacent Wake and Granville County (N.C.). The third school, Rogers-Herr, has the lowest Hispanic population (12%) and is somewhere ‘between’ E.K. Powe and Chewning demographically. It is located outside of Durham’s urban core in a somewhat more suburban setting which is less densely populated than the center of Durham but more populated than the rural setting of Chewning. Academically, Chewning was the lowest-performing school during the 2006-2007 academic year, with less than 50% of students overall performing at grade level on end-of-year standardized tests (NC Department of Public Instruction 2007). E.K. Powe had between 50% and 60% of students performing at grade level, and Rogers-Herr had the highest-performing students, with 60-80% of students performing at grade level.

4.3 Speakers

Speakers were drawn from the three schools in Durham (one elementary school and two middle schools) and ranged in age from nine to fourteen (grades four through eight). Speakers participated in sociolinguistic interviews performed by bilingual graduate student fieldworkers, mostly English (sociolinguistics) master’s degree
students associated with North Carolina Language and Life project (NCLLP) at North Carolina State University. These interviews were conducted during late 2006 and early 2007, and are currently archived online in the Sociolinguistic Archive and Analysis Project database (SLAAP) (Kendall 2007). In most cases, the Spanish and English interviews were conducted on different occasions by different interviewers. Speakers, especially in Durham, were interviewed in dyads whenever possible to increase comfort level. Topics in both the English and Spanish interviews included the typical everyday interests of U.S. elementary and middle school students: school events, movies, recreational activities, video games, and dating. Interviews at E.K. Powe took place in the classroom of the school’s ESL teacher. Middle school interviews were usually held in whatever empty classrooms were available during the school day, though a few interviews took place in an empty cafeteria or the school media center. Either immediately before or after the conversational interview, fieldworkers briefly collected survey data which listed basic demographic and social information for the speaker, including birthplace, LOR, family composition, best friends, music tastes, language use patterns (who spoke which language(s) to whom in the family), and language in which child learned to read.

In addition to English and Spanish sociolinguistic interviews, which typically lasted around 45 minutes and were conducted during the school day, structured education data were collected from some students. Though these measures were not used directly in this dissertation analysis, they were occasionally used to verify the
child’s reported literacy in English and/or Spanish. The academic language protocols included a reading/translation task in which children read 1-3 short picture book stories in English, then spontaneously translated the stories orally to Spanish, followed by standardized testing in English and Spanish, using the Woodcock-Johnson (WJ-III) tests of achievement (Schrank et al. 2005).

4.4 Data and coding procedures

4.4.1 Overview of coding methods for unmarked tense variable

Tokens of unmarked and marked past tense forms were extracted by listening to audio data which had been digitized and uploaded to SLAAP. Once a token was identified, it was transcribed in an Excel sheet and coded for overt tense marking, the dependent variable. The token was then furthered coded for the following ten independent variables, the first seven of which are linguistic factors and the last six of which are social factors (Table 4):

<table>
<thead>
<tr>
<th>FACTOR</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Verb Class</td>
<td>a. Suppletive (e.g., is/was; go/went)</td>
</tr>
<tr>
<td></td>
<td>b. Doubly marked (e.g., leave/left; say/said);</td>
</tr>
<tr>
<td></td>
<td>c. Internal vowel change, (e.g., come/came);</td>
</tr>
<tr>
<td></td>
<td>d. Change in final segment (e.g., have/had, bend, bent, try, tried);</td>
</tr>
<tr>
<td></td>
<td>e. Regular (-ed) consonant cluster verbs (e.g., talk, talked)</td>
</tr>
<tr>
<td></td>
<td>f. Weak syllabic (start/started) and</td>
</tr>
<tr>
<td></td>
<td>g. Nonverbal/lexical ending in consonant cluster (best, friend)</td>
</tr>
<tr>
<td>2. Phonological Environment</td>
<td>a. _V (ate and…)</td>
</tr>
<tr>
<td></td>
<td>b. _C (ate when…)</td>
</tr>
</tbody>
</table>

Table 4: Variables in quantitative analysis of unmarked tense
4.4.2 Variationist factors

Factors 1. and 2., verb class and phonological environment, were discussed in sections 1.2.1-1.2.2. in the context of Wolfram’s Vietnamese English studies and Bayley’s subsequent Chinese English studies. A few nominal changes were made to the methodologies used in those studies; for example, Wolfram coded speakers into four age groups spanning into adulthood and two LOR groups. While in earlier stages of analysis (Callahan 2008) I chose to group speakers into 3 LOR categories to facilitate Goldvarb analysis (as this program only admits categorical data), the logistic regression used in
this study permitted ages and LORs to be entered as continuous data (i.e. in numeric form).

Tokens that were excluded included negative phrases (*I didn’t know/she didn’t talk*), since in English tense is marked on the negative auxiliary and not the main verb. Contracted and zero copula forms (*There’s a girl in my class, They 0 so cool*) were excluded as well: though I did observe these forms in past tense environments and in the midst of past tense narratives, it was not possible in these cases to distinguish underlying present tense marking (*is/are*) from underlying past tense marking (*was/were*). One further addition to Wolfram’s coding categories was my decision to code for following phonological environment (*_C, _V or _P*) not only following cluster forms such as *kissed* or *leaned* (which would typically show a phonological effect of reduction preconsonantally), but also for forms which took a replacive final consonant (*agree/agreed, die/died*). The rationale here was that while Spanish phonotactics do permit postvocalic, word-final alveolar stops (*V+d* as in *verdad* ‘truth/true’, *pared* ‘wall’ and, less often, *V + t* in non-Spanish (including indigenous) borrowings, e.g. *yogurt, closet, Nayarit* (state in Mexico)), the */t, d*/ segment itself is reduced or deleted in Spanish in qualitatively different ways. For similar reasons, doubly marked forms (*keep/kept, hear/heard*) were coded for following phonological environment in cases which the internal vowel change often involved a contrast that which not exist in the Spanish vowel system (i.e. ‘keep’ and ‘kept’ exploit the non-native */i/* vs. */ɛ/* but not ‘tell’ vs. ‘told, in which */ɛ/* might be rendered as [e] but */o/* remains distinct). Finally, in the case of
weak syllabic forms (*needed, started*), I also coded for following phonological environment since unstressed syllables in English can be variably reduced or even deleted. When past tense morphology *could* be reduced or deleted in Spanish or English generally, I coded the phonological environment in order to have a record of possible effects.

### 4.4.3 Lexical/inherent aspect

Lexical aspect was coded as a property of the Verb Phrase (VP) as constituted by the main verb and its complements in the sentence, *e.g.* *get on my nerves* (Jesy, dps0970d), not simply *GET*. This distinction is necessary since the semantic variable may be coded different depending on the verb’s complements: thus, *get on my nerves* was classified as a (atelic/unbounded) activity with no inherent endpoint, whereas *get home* (Fely, dps1271d) was classified as a (telic/bounded) accomplishment. Next, the following tests were applied in order to classify the sentence into the four relevant categories (reprinted from Shirai & Andersen 1995: 749):

#### Step 1: State or nonstate

Does it have a habitual interpretation in simple present tense?
- If no -> State (*e.g.* *I love you*)
- If yes -> Nonstate (*e.g.* *I eat bread*) -> Go to Step 2

#### Step 2: Activity or nonactivity

Does ’X is Ving’ entail ’X has Ved’ without an iterative/habitual mean-ing? In other words, if you stop in the middle of Ving, have you done the act of V?
- If yes -> Activity (*e.g.* *run*)
- If no -> Nonactivity (*e.g.* *run a mile*) -> Go to Step 3

#### Step 3: Accomplishment or achievement

[If test (a) does not work, apply test (b), and possibly (c).]
- a) If ’X Ved in Y time (*e.g.* 10 minutes)’, then ’X was Ving during that time.’
If yes -> Accomplishment (e.g. *He painted a picture.*)  
If no -> Achievement (e.g. *He noticed a picture*)

b) Is there ambiguity with almost?  
If yes -> Accomplishment (e.g. *He almost painted a picture* has two readings: he almost started to paint a picture/he almost finished painting a picture.)  
If no -> Achievement (e.g. *He almost noticed a picture* has only one reading.)

c) 'X will VP in Y time (e.g. 10 minutes)' = 'X will VP after Y time.'  
If no -> Accomplishment (e.g. *He will paint a picture in an hour* is different from *He will paint a picture after an hour,* because the former can mean that he will spend an hour painting a picture, but the latter does not.)  
If yes -> Achievement (e.g. *He will start singing in two minutes* can have only one reading, which is the same as in *he will start singing after two minutes,* with no other reading possible.)

Finally, in order to assess the effect of telicity overall (i.e. whether or not the VP has an inherent, natural endpoint), states and activities were further coded as one category (telic predicates) and accomplishment and achievements were similarly coded into one category (atelic predicates).

### 4.4.4 Discourse structure

As noted in Chapter 1, some studies, beginning with Kumpf (1984), approached tense marking patterns in second language narratives by comparing unmarked forms with the use of the so-called ‘historical’ or ‘conversational present.’ Here, the use of the conversational present (CP) is explained in terms of a discourse function in which the verb form refers to events which are semantically past; however the speaker describes them in the present tense as a pragmatic device to lend a sense of ‘immediacy’ to his
narrative. The excerpt below gives an example of a narrative from high-proficiency speaker Julio (dps0402d, 254s):

Julio: In "America's Funniest Video," have y'all seen that show?

INT: Uh-huh

Julio: There was this man, he was tryin to- tryin to sell this cat. And he was, he was sayin like um, on camera, he was sayin that there wa- thi- that would be the nic- that was the nicest cat he's ever known. And then, when he takes him out of the bag, or the box, the box I mean, he bites him on the leg.

INT 1: <gasp>

INT 2: <laughs>

Julio: …and then he say he'd be the nicest cat in the world.

The two phrases in which Julio uses the CP (in “takes” and “bites”) are underlined in the dialogue above (‘when he takes him out of the...box...he bites him on the leg”). Julio’s use of the CP is typical in that he uses it to deliver the ‘punchline’ of his anecdote, which is simultaneously the climax of his oral narrative as a complicating action (Labov & Waletzky 1967). The joke is funny since it is ironic that the “nicest cat [the man] has ever known” would bite its owner in the cat’s video debut; here, the CP highlights the chunk of narrative Julio’s speakers must attend to in order to ‘get’ the joke. Though his story is initially framed as occurring in the past in an opening clause which has tense marked on its narrative ‘head’ (“There was this man...”), Julio exploits
explicit present tense morphology (3rd sg. -s) to move the most important part of his story forward. After it is clear that his joke has succeeded with the audience (“INT 1: <gasp>/INT 2: <laughs>”), Julio shifts back into using unmarked past tense “…and then he say he’d be the nicest cat in the world”).

As in some previous analyses of unmarked tense variation (Bayley 1994; Wolfram, Christian, and Hatfield 1986a) Julio’s principled, native speaker-like use of CP (with or without ‘tell-tale’ 3rd sg. marking) was not immediately observed for the beginning language learners in this study. At first glance, these speakers seem not to have (yet) acquired this particular discourse function. The lowest proficiency speakers did use alternate paralinguistic and linguistic cues like increased volume, intonation contours, and eye contact to mark narrative functions, however they did not seem to use CP, at least in the paradigmatic sense. The excerpt below features Enrique, a speaker who has lived in the U.S. less than 1 year, one of the lowest LORs in this study. This excerpt gives a sense of the character of tense alternations which occur the very beginning language learner narratives in the study. Enrique (dps049d/1580s) describes to the interviewer what happened on a recent field trip to Washington, D.C.:

Enrique: I **look** the airplane and…

INT: At the museum?

Enrique: Uh-huh, and…I **look** the animals…I **look** the animals…I **go** to… *fui a comprar cosas in una tienda* (‘I went to buy some things in a store’)

The interviewees in this study, especially the lower-proficiency speakers, were
made aware (in most English interviews) that if they didn’t know a word in English, they could say it in Spanish. Here, Enrique uses a code switch (“fui a comprar cosas in una tienda/’I went to buy some things in a store’”) as a strategy to move the events of his narrative forward and keep the attention of his audience (a bilingual fieldworker).

Initially, these paralinguistic cues seemed to correlate more closely with the function served by CP in native (MAE) speech.

However, in order to test formally for the patterning of conversational present, narratives from 15 of the low- to mid-proficiency speakers were transcribed and coded for the foregrounding vs. backgrounding function vs. tense unmarking. The first completed narrative to appear after the first five minutes of the interview was transcribed and coded. Here, for example, Alejo (dps049/345s), a mid- proficiency speaker, recounts what happened in a movie he recently saw:

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INT: What happened?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alejo: There’s a rat that— it’s in—I don’t remember— he’s in another state.</strong></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>And then he—another rat come for the—</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>can I say it in Spanish, that word?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From the taza... from the thing that—</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>INT: The spoon? The cup, I mean, the cup?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alejo: no, where you pee</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>INT: Oh, the toilet?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alejo: Yes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>INT: Uh-huh</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alejo: A rat come from over there</strong></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>INT: Uh-huh</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alejo: and the rat that was in the other state—</strong></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>he say ”you need to get out of here.”</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>And he say ’No.”</strong></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
“Do you have TV?” And he say “yes.”

And then the rat that come from the toilet say, “Oh my God, look at this monster!”

That was a big TV.

And then, he say “Do you wanna...bañarte en el jacuzzi?”

INT: <laughs> Uh-huh

Alejo: ...and that was the toilet.

And he say “Sure.”

And then he say “OK, go down. Go down.”

And he say “OK, here I’m going... One, two— wait!”

And then he drove the other rat to the water

and he go down, down

and there was a big city, like big city.

INT: Under the water?

Alejo: Yes. The rats made a city.

And there was a--

The rat was a girl.

And another rats want to kill her. Another rats.

And he have a diamond.

That’s why they want to kill her.

But she want to escape.

And the guy that come down

he say

he want to get to hiselfs [sic] again

And the rat say—

the girl say, “I’m sorry. I have m— problems.”

And the rat say “Please, please. I have a lot of diamonds up there.”

And... I don’t remember.

The clauses which form Alejo’s narrative are underlined. The verbs in bold were considered candidates for the unmarking analysis. When it could not be determined if the tense was present or past tense (There’s, It’s) or if the clause/verb form did not form part of the narrative (e.g. where you pee, as an ‘aside’ when Alejo asks for clarification), the token was excluded from the analysis. In addition, if a clause was a false start (And there was a—, he say—) it was excluded. A clause which ‘pushed the event line forward’
(Kumpf 1984: 141) was coded as a foreground clause; correspondingly; background clauses were defined as those which ‘set the scene, make digressions, change the normal sequence of events, or give evaluative remarks.’ (1984: 141). These procedures for coding foreground vs. background clauses are identical to the methods used in the 1980s Vietnamese English tense unmarking studies which examined the use of CP (Wolfram & Hatfield 1984; Wolfram & Hatfield 1985).

4.4.5 Other exclusions

I excluded tokens from the analysis if they seemed to be lexicalized, including tokens ‘and’ and ‘just.’ I initially noticed very few full clusters at the ends of these words, therefore, during coding of the first 31 speakers, I coded the first 5 instances of each token and spot-checked the remainder of each interview. I did not transcribe a single unreduced form in this subsample of 155 cases; therefore I hypothesized that these entries may not be variable for many speakers (i.e. they are acquired without the final segment).

4.4.6 Statistical methods

Initially, basic graphs (e.g. scatterplots; raw incidence of verb marking by speaker, interview, or age) and descriptive statistics (e.g. cross tabulation of phonological environment vs. marking; verb frequency histograms) were used to visualize data and check for obvious interactions among factors. This initial, exploratory phase of analysis was carried out using the same program in which the data was tabulated and coded, Microsoft ® Excel ® version 2010 (version 14.0 for Windows).
Second, a logistic regression using the software program SPSS Statistics (SPSS 21 for Windows) was used to comprehensively investigate the constraints on tense marking patterns in terms of the independent factors 1.-12., listed in Table 1, e.g. as measured by standard correlation statistics (e.g. Pearson’s $r$ and corresponding significance level $p$). The initial regression coded linguistic factors. 1-6. (verb class, phonological environment, lexical aspect, embedded or auxiliary status of verb, frequency class) and social factors 8., 9., 10., and 12. (interview, interviewer(s), gender, and school) as nominal/categorical variables. Only two factors, 7. (Length of Residency) and 11. (school) were entered as continuous variables.

Subsequent analyses were planned, as needed, with re-codings of the independent factors as entered. For example, in subsequent runs, frequency was reconsidered a continuous measure vs. a categorical one (incidence proportion vs. categories high/medium/low)\textsuperscript{3} in order to identify the best-fitting model. These decisions and their rationales are discussed in detail in section 5.3.5. Finally, in some cases where the graphing functionality or SPSS was limited, the analyses were run in the statistical environment R (Version 2.15.1 for Windows) using package ‘nlme,’ designed for fitting linear mixed-effects models.

\textsuperscript{3} Erker and Guy (2012), in a study of variable subject personal pronoun (SPP) frequency in Puerto Rican Spanish, found that frequency was best coded as a discrete vs. continuous variable since it acted as ‘gatekeeper’ on stable patterns of variation in other factors. Common verbs which occurred above a certain frequency (e.g. creer ‘to believe’, saber ‘to know’, and ver ‘to see’) seemed to amplify the effects of other factors like person and number. Below a certain threshold, however, these effects rapidly diminished (person and number showed no significant effect on SPP patterns for verbs amenzar ‘to threaten’ or convivir, ‘to exist’). We will return to a discussion of frequency in Durham HE in section 5.3.5.
Given this overview of the Durham site, speech community, and speakers (sections 4.1-4.3) as well as documentation of coding procedures for both the linguistic and social categories (4.4.1-4.4.5) and description of statistical methods (4.4.6), we will move on to section 5., which presents the results of the study.
5. Results

5.1 Introduction

In this section, I will present the results of the statistical analyses described in 2.3.6 in both descriptive terms (by laying out the patterns present in the data) as well as in predictive terms (by accounting for correlations in the dependent variable and sociolinguistic factors, including co-functioning of constraints at multiple linguistic levels). In 5.2, I will comment on overall distribution of marked tokens for tense in the data. In 4.3, I will present the overall results of logistic regressions analyses run on each of four linguistic or internal factors in the model (phonological environment, verb class, lexical aspect, and frequency), and provide guidelines for interpreting these statistics. After this overall preview of the results of the internal factors, I will examine in detail each linguistic factor individually in sections 4.3.1-4.3.4. Next, in section 5.4, I will discuss the influence of discourse/narrative context on unmarking patterns by presenting the results of a subsample of the data which was coded for foregrounding vs. backgrounding function. Finally, in section Results by social factor group, I will present the results for each of the social factors hypothesized to correlate with unmarking (length of residency (LOR), age, school, gender, country of origin, literacy in English/Spanish, and gang affiliation) and discuss in detail the effect of LOR in relation to tense unmarking.
5.2 Overall distribution of tokens

Of the 2012 total nominal and verbal tokens (i.e. marked and unmarked) produced by the all speakers in the study, the mean number of tokens produced per speaker was 46 (median= 36), with the speakers producing the fewest amount of tokens (Anabel and George, each with a LOR= <1 year) producing only one token each and the speaker producing these most tokens (Marisa) producing a maximum of 164 tokens. The spread for quantity of tokens produced was fairly high (standard deviation= 40.5), though, interestingly, there was a low correlation (r=.23) between length of residency and number of tokens produced: in other words, low LOR didn’t necessarily keep speakers from talking. In fact, of the 4 speakers who produced over 100 tokens each, one had been in the U.S. only 2 years (Marisa; n=164) and the other for 4 years (Fely; n=106). Figure 9 lists percentage of unmarked tense for all 44 speakers in the study, with names of speakers below for reference. As we might expect, there is a wide spread, with some speakers (e.g. Cleo, LOR=10; Selena; LOR=2) marking a relatively low number of tokens and other speakers (Tony, LOR=16) marking almost all of his tokens.
5.3 Results by linguistic factor group

Table 5 shows the results of logistic regressions run on each of the main linguistic factors in the model, demonstrating that all four (phonological environment, verb class, lexical aspect, and frequency) correlate significantly with rates of past tense marking.
### Table 5: Logistic regression results for linguistic (internal) factors

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\chi^2$ (Model)</th>
<th>$-2\log$ Likelihood</th>
<th>Sig.</th>
<th>Pseudo $R^2$ (Nagelkerke)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Phonological Environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Consonant</td>
<td>-0.515</td>
<td>28.6</td>
<td>1272.951</td>
<td>p&lt;.000</td>
<td>0.04</td>
</tr>
<tr>
<td>b. Pause</td>
<td>-1.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowel [ref]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Verb Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Cluster verb ('picked')</td>
<td>-0.520</td>
<td>441.6</td>
<td>2097.629</td>
<td>p&lt;.000</td>
<td>0.28</td>
</tr>
<tr>
<td>b. Final Repladive C ('had')</td>
<td>0.676</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Doubly marked ('told')</td>
<td>0.924</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Vowel Change ('came')</td>
<td>1.102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Supplative ('want')</td>
<td>2.179</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak Syllabic ('wanted') [ref]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Lexical Aspect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Accomplishment</td>
<td></td>
<td>335.5</td>
<td>2298.175</td>
<td>p&lt;.000</td>
<td>0.21</td>
</tr>
<tr>
<td>b. Achievement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. State</td>
<td>1.188</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity (reference)</td>
<td>1.706</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Frequency</td>
<td>0.010</td>
<td>355.7</td>
<td>2268.103</td>
<td>p&lt;.000</td>
<td>0.22</td>
</tr>
<tr>
<td>5. IOR</td>
<td>0.098</td>
<td>187.5</td>
<td>2566.011</td>
<td>p&lt;.000</td>
<td>0.05</td>
</tr>
</tbody>
</table>

#### 5.3.1 Interpretation of logistic regression statistics

Before investigating these initial results further, it may be useful to provide a brief explanation of the meaning of each statistic given in Table 5 since researchers across fields (SLA, sociolinguistics, etc.) may have varying degrees of familiarity with statistical analysis of linguistic data. In broad terms, the individual figures shown in Table 5 demonstrate, in general, whether the relationships observed between each factor and the incidence of unmarked tense are due to chance.
First, the Chi-Square statistic (χ²), used frequently in studies of categorical/structural variation, provides a measure of the likelihood that each factor (phonological environment, verb class, etc.) functions independently from the occurrence of unmarked tense in this dataset: it describes ‘goodness of fit’ in terms of a comparison between the distribution we would expect if the variables were entirely independent from the phenomenon under investigation and the actual data distribution which we observe (in terms of the number of unmarked vs. marked forms which respect to each factor above, 1.-4.). Here, the greater the Chi-square statistic, the greater the likelihood that the variables (e.g. phonological environment vs. marking patterns) are actually dependent in their functioning—and the less likely that the null hypothesis, the assumption of no effect, or random variation, is correct.

Ultimately, we can use the Chi-Square statistic, in conjunction with the number of degrees of freedom (a measure of independence of the various factors: from a data table, we would calculate (R-1)*(C-1) where R=number of rows and C=number of columns) to consult a standard normal distribution table in order to look up an approximate probability, or p-value. Here, a value of .052 would mean, more or less, that if the data distribution is in fact simply due to chance, one would have (roughly) a 5.2% chance of finding a result as extreme as the one in a given data set. In other words, the null hypothesis (the functioning of random chance) is about 5.2% likely to be correct. It follows that a lower p-value can point to the likelihood of a stronger relationship among the factors.
The -2 log likelihood figure also compares the fit of two models, one of which assumes the null hypothesis is correct and the other which follows the study hypothesis (e.g. ‘phonological environment correlates in a statistically significant way with unmarked tense’). Similar to a Chi-square statistic, the -2 log likelihood is an index of how unlikely it is that a particular arrangement of data might be produced simply due to chance (i.e. in a universe where none of the hypothesized constraints actually predicted the variation under investigation).

The final statistic, a Psuedo R-squared ($R^2$), is not widely used in studies of (categorical) variation, but is useful here since the traditional R-squared itself is widely used and recognized. Also known as the ‘coefficient of determination,’ a traditional R-squared value measures, in general, the proportion of variance accounted for by a researcher’s model. Typically, R-squared ranges from 0 to 1.0, with values closer to 1 indicating a better fit of the model to the variation in the data (we could say, in turn that an R-squared value of .12 accounts, roughly, for about 12% of the variation present in the data). It is important to note here that this traditional R-squared measure is used only with continuous variation (e.g. fundamental frequency, which might range in a given sociolinguistic study, from 100Hz to 200Hz), not categorical variation where, for example, only two discrete values are possible (0=unmarked/1=marked with nothing in between). However, in order to test goodness of fit in a general way for models which attempt to predict categorical variation, statisticians have developed ‘psuedo’-R-squared

---

1 The -2 Log Likelihood is similar to an F-test in linear regression.
measures which use a similar scale, from 0 to 1, with higher values indicating better model fit; this measure is especially useful in comparing effects within one model (i.e. pseudo-R-squared values can be meaningfully compared with each other within one analysis, but not as readily across alternate models with different sets of factors (Hosmer and Lemeshow 2000).

5.3.2 Phonological environment

Figure 10 shows raw counts of the phonological variable of consonant cluster reduction (CCR), or –t/-d absence, on verbal tokens which are candidates for the unmarking process (that is, as discussed in sections 3.3 and 4.4.2, all case of past tense that were potentially formed through the addition of /t/ or /d/ that might result in a consonant cluster (e.g. missed) as well as monomorphemes whose final syllable ended in a cluster (mist)).
Figure 10: Raw counts of –t/-d absence by phonological environment

Setting the Y axis equal to count vs. percentage allows us to see the raw totals of kinds of verbs, marked and unmarked in the data, relative to each other. The pattern observed here reflects the trend which has been widely demonstrated in studies of CCR in both native and learner English speech communities (Fasold 1972, Guy 1980, Labov 1972a, Santa Ana 1991, Wolfram 1969, 1974): pre-consonantal tokens favor the process of reduction (a token like ‘picked through’ is reduced most often) and prevocalic tokens disfavor it (‘picked up’ is reduced least often). Rates for pre-pausal tokens (‘the one I picked’), however, which usually fall between the two levels, are lower than either preconsonantal or prevocalic tokens. For now, we will lay aside this effect and return to
the discussion later vis à vis similar effects in local varieties of African American English later, in section 6.1.2.

5.3.3 Lexical Aspect

Figure 11 shows the results for past tense marking and lexical aspect based on the definitions of these different categories proposed in section 4.4.3.

Figure 11: Raw Counts of surface past tense marking by lexical aspect

First, we can see that there are more marked tokens overall than unmarked: most verbs in the data set are marked with Standard English type past tense marking. Second, the aspect variable is marked at the relative rates we would expect, with telic predicates...
marked overall relatively more often than atelic predicates--with the exception of states, which are apparently marked at rates twice as high as the trailing category, accomplishments. This is a surprising result since we would expect that statives would be highly unmarked if lexical aspect does in fact influence marking patterns. However, there is evidence that this effect appears to be at least partially lexically-based: one token type, the copula, accounts for almost 17% of the total Ns in the data: copula is the most frequent token type overall. Furthermore, of the 675 tokens which describe a state (in addition to copula, verbs like ‘have,’ ‘want,’ and ‘believe’), a full 359, or 53% are copula tokens. Of these copula tokens, an overwhelming majority, or 97% are marked--thus, copula is almost categorically marked and makes up over half of the stative tokens.

These high rates of copula marking (along with high incidence of copula itself) have also been found in previous studies of English interlanguage variation: Kumpf (1984), in the study described in section 2.3.1, found that over half her stative tokens are copula, and that, furthermore, copula is tensed 100% of the time. Ellis (1987), in an analysis of style-shifting among 17 EFL University students, found that that copula marking showed intermediate effects between that of regular verbs and irregulars. When Ellis’ subjects, (low-intermediate learners from mixed language backgrounds), were given extra planning time to orally narrate the events they saw in a sequence of pictures, the planned (spoken) narratives showed a higher incidence of copula marking,
from 60 to 75% ($\chi^2 = 4.36; p< .05$).\(^2\) Bayley (1994) does not report especially high levels of copula marking in his Chinese English studies, but does justify bifurcating copula into two different factor groups in his Varbrul (Rousseau & Sankoff 1978, Sankoff 1988) analysis, since all forms of copula except for 1\(^{st}\) singular pattern with internal vowel change forms over suppletives (i.e. ‘is’ $\rightarrow$ ‘was’ is more similar to ‘come’ $\rightarrow$ ‘came’ than, say, ‘go’ $\rightarrow$ ‘went’).\(^3\) In any case, there is sufficient evidence to consider copula as a class (or classes) unto itself in studies of English interlanguage variation. Indeed, as shown in Figure 12, when state/copula tokens are excluded, marking patterns at all levels of proficiency follow the prediction for marking by lexical aspect, where activities are marked less often than achievements, though the lowest LOR speakers mark significantly fewer verbal tokens overall.

---

\(^2\) Though acquisition is assumed to differ in qualitative terms across learning environments (e.g. naturalistic acquisition in the L2 speech community vs. formal instruction/classroom learning), we might also speculate that more or less attention to speech could outline the same trajectory as acquisition over time.

\(^3\) Bayley (personal communication) reports that the relatively low rate of copula marking in his Chinese English data may be indicative of an L1 effect since copula is widely non-obligatory in Chinese: among lower proficiency learners, he found copula could be absent, while, among higher proficiency learners copula could be present but not marked.
5.3.4 Verb class

The saliency hierarchy, represented by verb class, by contrast, does follow the expected rank order observed in previous studies, with a reverse in doubly marked and internal vowel change type verbs (Figure 13).
Figure 13: Raw counts of past tense marking by verb class

Here, the regular verbs—weak syllabic ‘wanted’ and regular cluster ‘picked’ and ‘leaned’—share relatively more segments with their stem, making them less phonetically salient to learners⁴. By contrast, doubly marked forms like ‘tell → told’ and suppletive forms like ‘go → went’, show markedly higher rates of marking, though they also represent the majority of tokens overall.

⁴ The effects of stress may be relevant as well; for example, internal vowel change tokens may share relatively more segments than weak syllabics, however the vowel in IVC forms is always stressed (be’come → be’camed) whereas the –Id form never is (‘want → ‘wanted).
Furthermore, overall rates of marking by verb class show sensitivity to phonetic salience at all proficiency levels, from speakers who have lived in the U.S. for 1-3 years (LOR 1) to speakers who have lived in the U.S. for 6 or more (LOR 3) Figure 14).

![Graph showing percentage of past tense marking by verb class across different periods of residence.]

**Figure 14: Percentage of past tense marking by verb class**

When the results are articulated by proficiency it becomes evident that the very beginning language learners (LOR 1: 1-3 years) do narrowly follow the saliency hierarchy as-is, and the reversal in doubly marked and internal vowel change verbs happens (though not at dramatic levels) in the latter two LOR periods. However, overall we can identify two main trends: 1. marking remains lower overall at the lower proficiency levels and higher overall at the higher proficiency levels, regardless of verb
class and 2. rates of marking by verb class tend to stabilize at the higher proficiency levels, though sensitivity to the saliency hierarchy remains.

To round out a discussion of verb class, we will consider a comparison of three previous studies of variation in past tense marking in learner English with the results from this study. Here, Figure 15 shows the results of the present study (Durham Hispanic English) alongside Wolfram and Hatfield’s (1986) adult Vietnamese English speakers, Bayley’s (1991) young adult Chinese English speakers, and Adamson’s (2009) elementary-aged Chinese English speakers.

Figure 15: Percentage of unmarking by verb class and proficiency for 3 studies of English interlanguage
As mentioned in section 5.3.2, Bayley ultimately combines into classes his 1a. strong verbs (vowel change: ‘come’ $\rightarrow$ ‘came’) and 1b. copula other than 1st singular (‘is’ ‘was’) as well as 2a. replacives (‘send’ $\rightarrow$ ‘sent’) and 2b. weak nonsyllabics (‘walk’ $\rightarrow$ ‘walked’), accounting for the blank cells in the table below the results. Wolfram and Hatfield, similarly, combine regular verbs into one class (cluster verbs and weak syllabics), and Adamson only considers three categories in his study (vowel change, replacives, and suppletives).

Despite the differing methodological conventions, however, the graph illustrates the striking effect of phonetic saliency across language learning communities and proficiency levels. The main difference which can be observed for Durham HE is the low overall rates of unmarking by the high proficiency speakers; as a group, these speakers mark many more verbal tokens than their language-learning cohorts in the other 3 studies (staying below 50% unmarking for all verb classes except for cluster verbs). One fundamental difference in the overall profile of each community in the comparison is that the Durham speech community includes both recent immigrants as well as native speakers whereas Wolfram and Hatfield’s high proficiency category topped out at 7 years, Bayley’s highest proficiency speaker (Guo Chang) had been in the U.S. only 54 months, and Adamson’s longitudinal study included children in their first 3 years of residence in the U.S.

It is worth noting, however, that the native/non-native distinction in Durham, which may in general seem very primary, does not play out in clear-cut terms.
Linguistically speaking, of the 11 U.S.-born speakers in the study, (Lou, Lalo, Pily, Jesús, Federico, Jorge, Rolando, Lori, Micha, Michelina, Alejo), all used the unmarking process in at least one non-cluster verb (i.e. outside of environments where the process could simply be attributed to phonological reduction). Lori, who is 12 years old, was born in North Carolina, left at age 5 to live in the Mexico City, and returned to the U.S. at age 7 (when, she says, she “forgot all her English”). Alejo, 9, was born in Pennsylvania where his mother had family, attended first through fourth grade in Michoacán (Mexico), and had only recently returned to the U.S. at the time of the study. Lou, one of the oldest speakers in the study at 16, was born in Waco, TX, lived across the border in Mexico until he was 1 year old, then moved with his family to Rougemont, N.C.

The native/nonnative distinction also does not play out as one might think in social space (in terms of ties to home and vacations abroad). In my experience as a teacher, it was often harder for the foreign-born, undocumented students to travel with their families outside the U.S. since they would have to risk another perilous ‘crossing’ to re-enter. Thus, some students who were born in the U.S. actually maintained closer ties with Spanish-speaking family outside the U.S. and were able to visit them more frequently and freely. Moreover, in the every day, the U.S.-born and non-U.S. born students spent much of their daily schedule together: all 11 speakers attended regular ESL classes with their lower proficiency peers, whether in a stand-alone class (a period during the middle school day) or during a ‘pull out’ session with their elementary school ESL teacher. Thus, the lines between ‘native’ and ‘nonnative’ speaker of English
were often blurred in both linguistic terms (successive rates of unmarking) as well as in social (institutional, political) terms: policy and pedagogy applications will be discussed in Chapter 5.

5.3.5 Frequency

Much of the recent ‘usage-’ or ‘exemplar-based’ (Bybee 2001, 2002c) work on the role of frequency in linguistic variation has been done on phonological structures like variable /s/-reduction in Spanish (Brown & Torres Cacoullos 2002), variable coronal stop deletion in English (Guy et. al. 2008), or t-to-r in English (Clark and Watson 2011). Erker and Guy (2012), in a frequency-based study of variable subject personal pronoun expression in Spanish, present a number of distinct issues at stake in quantifying usage-based effects in studies of syntactic variation. The data in this study follow the first (typical) trend they describe: there are many highly frequent tokens (‘be,’ ‘friend’) and, conversely, few very rare tokens (‘betray,’ ‘scientist’), presenting issues for a statistical analysis which is very sensitive to sample size. Other issues include whether to consider verbs as abstract units or parts of collocations with their objects (‘get X’ vs. collocations like ‘get _mad’ vs. ‘get_through’ etc.), whether to use local or global corpora, whether to consider frequency as a continuous vs. categorical constraint, and whether to consider frequency as a traditional ‘factor group’ at all, or one that may intersect in meaningful ways with independently functioning sociolinguistic constraints. Finally, there is the question of whether frequency resists or accelerates change, as with morphological generalization. Poplack (1992), for example, showed that for variation in the subjunctive in Canadian
French, very high frequency verbs like *faîloir* preserved older forms. Similarly, the well-known pattern in the history of English preterit is a trend from strong forms (‘dreamt,’ ‘dove’) to weak forms (‘dreamed,’ ‘dived’) based on high token frequency and low token type.

In this study, I followed Erker and Guy (2012) in coding all data by raw frequency count (a token which occurred 15 times received a frequency rating of 15) and designating high frequency tokens as ones that occurred in the top 1% of all verbal tokens. In our data set of 2012 tokens, this 1% threshold occurred at n=20 (that is, if a token occurred 20 times or more, it was categorized as a high-frequency token). For phrasal verbs (‘get ahead’ vs. ‘get through’) I coded the verb and its complement. Figure 16 illustrates the distribution of all verbs in the data with a reference line at n=20, the cut-off point designated for high vs. low frequency.
As expected, learners mark high frequency tokens (like ‘die’, ‘win’ and ‘look’) at higher rates at all proficiency levels as compared to low-frequency items (like ‘blame’ or ‘switch’ or ‘cuss’).
However, this effect is not so straightforward. Figure 18 demonstrates how frequency itself is articulated by verb class.
Figure 18: Verb class by Frequency and LOR
The top half of Figure 18 illustrates that at all levels of proficiency, suppletives, the most commonly verbs used overall, are employed consistently by speakers at all LOR points at rates over three times that of the other verb classes. As the most phonetically distinct type of verb, indeed, they must be frequent to resist the forces of leveling or generalization over time: in Bybee’s (1995) terms, they have increased lexical strength. Bybee also notes that highly frequent, or entrenched words and phrases tend to be cognitively stored unanalyzed—which, I would argue, correspondingly makes them easy to learn and access for language learners.

5.4 Results by discourse factor group

A subsample of 12 speakers was chosen for the analysis of unmarked tense by discourse/narrative structure factor in order to test the hypothesis described in sections 2.3 and 4.4.4, that past marking is sensitive to foregrounding vs. backgrounding function. The first narrative of each of the speakers’ interview which contained 5 or more verbal tokens in obligatory past tense contexts (n=191) was transcribed and coded according to the procedures outlined in 4.4.4. The results are shown in Figure 19.
Figure 19: Percentage of past tense marking by speaker and narrative function (foreground vs. background)

Though this sample is limited, the overall trend is toward more marking in background clauses ($\chi^2 = 12.28; p<.05$): all but two speakers, Jorge and Alejo, categorically marked background clauses. However, the frequency effect described in 5.3.5 was identified for this variable as well: of the background clauses, over 80% (58/72) were
copula tokens (only 6/116 foreground tokens were copula tokens, and all were marked)

Figure 20: Background forms and marking
Figure 21: Foreground forms and marking

A few examples of the pattern, where copula tokens are used in backgrounds and are marked at high levels, are given below:

1. Pily/091d: One time I got really scared (.) it was in El Salvador (.) I was at the park...
2. Jorge/dps041d: Last time I visit (.) was when I was five (.) I want to go visit—a doctor...
3. Jorge Y./dps0550d: [Describing a video game he played with a friend] the mission was to—I don’t remember but it was something to do with hippies...
4. Marcela/dps058: We went like to a yard sale (.) there was—like toys (.) and we buyed it for him [her dog]...

---

5 Pauses of more than approximately one second are annotated (.); pauses of a second or less are annotated with –.
Copula is used the majority of the time for background functions like ‘setting the stage’ for the events in a story line, as in 1.-4. above. Adamson (2009) observed a similar effect in his Chinese English data, where by far the most common background verb was the copula. He notes that since background clauses are not absolutely required in narrative discourse for young speakers (Aksu-Koc and von Stuttenheim 1994 cited in Adamson 2009), the speaker has the freedom to use familiar verbs, whereas in the foreground they are constrained to use particular verbs which describe the events in a particular storyline. In our subsample, there may be some limited evidence to support this account, since Micha and Alejo, both beginning language learners (LOR=1), also had the lowest rates of foreground marking, however the sample is too limited to draw any definitive conclusions. If anything, there is evidence for copula-heavy backgrounding and marking-heavy copulas in the context of more marking overall in the background for all verbs.

5.5 Results by social factor group

Not unexpectedly, there were highly significant correlations between many of the social factors selected as independent variables, most prominently with LOR. Specifically, collinearity occurred with LOR and the following factors, age ($r=.50; \rho<.000$), gender ($\chi^2=1034.2; \rho<.000$), country of birth ($\chi^2=2326.3; \rho<.000$), language in which learned to read ($\chi^2=3433.2; \rho<.000$), gang affiliation ($\chi^2=810.6; \rho<.000$) and school

---

*All confidence intervals were set to 95%. For continuous variables (LOR, age) correlations are in terms of Pearson $r$ in terms of a paired-samples $t$-test.*
(χ² = 1641.3; p < .000). Highly significant correlations also occurred with the category gender itself, including age (χ² = 713.921; p < .000), language in which earned to read (χ² = 2198.1; p < .000), and country of birth (χ² = 2040; p < .000), and gang affiliation (χ² = 2326.16; p < .000).

Ultimately, however, there was no need to investigate further the relationships between social factors themselves, since a logistic regression with all social factors included (5.-10. in Table 4) demonstrated that only one showed a significant relationship with the occurrence of unmarked tense, LOR (p < .001)\(^7\). Figure 22 shows the general downward trend of the unmarking effect with respect to LOR.

---
\(^7\) Interestingly, when the phonological process is included in the model (i.e. accounting not only for the results of bimorphemic/verbal tokens like ‘pick’ but monomorphemic tokens like ‘test’), the relationship between reduction (whether verbal or not) and the following social variables does reach significant levels: age (p < .01), interview (p < .01) and, almost, country of birth (p = .051). Especially in light of the dependency between LOR and those social variables and the lower indices of significance within this model (i.e. more room for chance), this effect is not discussed here.
Figure 22: Percentage of past tense marking by speaker LOR

Figure 22 demonstrates that speakers are most likely to unmark their verbs during their first two years living in the United States; during this period, speakers will produce a zero-marked past tense verb nearly 6 times out of every 10. After this initial two-year period, rates of unmarking plummet to a low of 22% by LOR 4. The sharp falls at LOR 7 and 12 may not be representative of the overall trend since those time points represent limited numbers of speakers: only two speakers, Marcela and Rosa, represent LOR 7 and, together, only produce 12 tokens, the lowest quantity of any LOR time point. Similarly, LOR 12 is represented by one speaker, Rolando, who produces 22 tokens, and LOR 8 is represented by only four speakers (Marcos, Goyo, Rosa, and Jackson) who produce only 82 tokens. Thus, the LOR cells (LOR 7, 12, and 8) which exhibit sharp up- or down- trends are in fact the least representative of all the time points in the data.
5.6 Summary of results

In sum, emerging Hispanic English in Durham, N.C. shows a complex array of internal/linguistic, discourse-based, and social constraints which not only reliably constrain past tense marking patterns, but also constrain each other as part of a unitary, context-driven process of language learning. In large part, the same kinds of linguistic constraints and constraint hierarchies demonstrated for previous studies of English interlanguage, including Wolfram’s (1985) study of Vietnamese English and Bayley’s (1991) studies of Chinese English, are reflected in the Durham Hispanic English data. Taken together, this cohort of variationist studies of English interlanguage confirm the central influence of verb class on unmarking patterns in terms of a hierarchy of phonetic saliency, where tokens which are highly phonetically differentiated to learners (‘go’ vs. ‘went’) are marked at higher rates over tokens are not as phonetically distinct in terms of segments shared with a stem (‘learn’ vs. ‘learned’). The verb class constraint is reliably articulated by LOR and Frequency with the exception of copula tokens, which interact in meaningful ways with the verbal predicate state, the verb class suppletive, and the frequency constraint. Here, the data show that high frequency is simultaneously phonetically salient: the two effects cannot be divorced since they mutually support each other and, in turn, the process of language learning, which is largely driven by redundancy. The most perceptually salient tokens, irregulars, can only afford to occur at high frequencies to resist leveling by generalization, a process which has been spelled out over the long haul by historical change in English (strong to weak preterit drift). In
turn, learners reach for these familiar, distinct forms to structure their discourse in terms of low-risk, high-reward strategies like exploiting copula to fill in the background of a narrative in order to ‘set the stage’ for optionally-marked foregrounded events. Thus, on the one hand, in future studies, care should be taken to allow for the effects of super high-frequency tokens like ‘be’ and ‘go,’ which may skew the results of a traditional, variationist analysis (i.e. if factor groups are not specified at the word-level). On the other hand, in the context of an evidence-rich learning environment, learners may well be exploiting the synergistic effect exhibited by highly frequent tokens which are simultaneously highly phonetically salient.
6. HE in the community context

In section 3.1, I discussed how the phonological process of CCR correlates with particular patterns of social category membership in specific communities—for example, the variable shows uniform constraints (e.g. morphemic status vs. phonological environment) which occur in different rank orders, a pattern which has been reliably replicated in many studies over the past four decades. In this section I will return to a discussion of CCR in both in Durham HE and AAE in order to explore patterns of social accommodation vs. differentiation in this emerging dialect. Beyond its scholarly value as firsthand evidence of dialect emergence, a clear understanding of the operation of phonological and grammatical variables (CCR in conjunction with past tense unmarking) has real-world implications. These linguistic and social processes have direct effects on the school experience of developing HE speakers, such as in the context of standardized testing designed with a Standard English bias. Thus, section 0 will provide a description of the convergence and divergence of HE in terms of linguistic constraints which have been typically studied for native varieties of English. Section 6.1.2.1 will discuss relevant pedagogical applications of the general findings for both CCR and unmarked tense. Finally, section 7 will provide directions for future research and general conclusions for the study as a whole.
6.1 **HE and AAE in contact**

6.1.1 **AAE contact features**

Before discussing the quantitative results for CCR, I will provide a sample of some of the AAE features which were observed in the speech of HE speakers in the study (Table 6). These examples occurred during the course of transcribing and coding the data. Thus, there was not a focus on the unmarked tense variable; the goal was simply to document examples of AAE structures when they occurred to offer supportive data for the contact influence from this variety. The list is based on the canonical inventories of AAE features found in works such as Labov et al. (1968), Fasold and Wolfram (1970), Rickford (1999), and Green (2002), among others.

**Table 6: AAE Features in Durham HE**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Speaker/Interview</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Possessive pronoun double marking</td>
<td>Luisa/dps0280d:</td>
<td>Luisa: We made this friendship bracelet, but I forgot <strong>mines</strong>.</td>
</tr>
<tr>
<td>2. Habitual 'be' (plus lexical 'playin')</td>
<td>Lalo/dps 0730d</td>
<td>Lalo: [<strong>She be doin</strong> a] lot of things to me. A lot of — private things too.</td>
</tr>
<tr>
<td></td>
<td>Goyo/dps0080d</td>
<td>Goyo: my brother, he drives. He be—he be—he be—<strong>we be drivin</strong> over here.</td>
</tr>
<tr>
<td>3. demonstrative them + plural noun</td>
<td>Goyo/dps0080d</td>
<td>Goyo: and you know, they have <strong>them bumps</strong> in Mexico—like in some streets?</td>
</tr>
<tr>
<td></td>
<td>Che/dps1080d</td>
<td>Che: [talking about his uncle getting a prize at the Fair] He had one of them—<strong>them</strong>—<strong>whatchamacallit</strong> things.</td>
</tr>
<tr>
<td>4. a' + V-initial noun</td>
<td>Fely/dps1271</td>
<td>Fely: …because, um, a <strong>evil man</strong>, I don’t remember his name, he was trying to become king.</td>
</tr>
</tbody>
</table>

1 For comprehensive features lists for AAE, see Green (2002) and Wolfram and Schilling-Estes (1998). These sources were also used as reference in compiling Table 6.
The speakers who appear in Table 6 represent a cross-section of speakers in the study: Lalo and Goyo, for example, are both 13-year old gang-affiliated student at
Rogers-Herr Middle School. Lalo is of Salvadoran descent, was born in Los Angeles, and moved to Durham when he was 8 years old. Goyo arrived in the U.S. from Mexico when he was seven. Luisa, by contrast, is a nine year old 4th grader who attends E.K. Powe (no gang affiliations were reported for any of the elementary school students). She was born in Mexico, learned to read in Spanish, and came to the U.S. when she was seven. She listens to pop music and has a European American best friend, Alex, with whom she speaks only English. Fely also attends E.K. Powe. She is an eight year old third grader who arrived from Mexico before she started kindergarten and learned to read in English. Che is a 12-year old non-gang-affiliated Chewning student from Morélia, Mexico, who came to the U.S. even younger, when he was two years old, and also learned to read in English.

Several of the features above may not originate exclusively in AAE; a case could be made that at least some have their origins in Southern White Vernacular English (e.g. 3., 4., 8. and 9.). The relationship between AA and White varieties of English in the South is far from straightforward as variation may show both superficial convergence, functional divergence, quantitative vs. qualitative differences, and bilateral contact (AAE \( \rightarrow \) SWVE as well as SWVE \( \rightarrow \) AAE). The evidence presented in Table 6, however, demonstrates that key forms do pattern structurally and functionally with AAE, and several of them (e.g. habitual be, 0 possessive, etc.) are unique to AAE among American English language varieties. For example, when SWVE shows copula deletion, it tends to
occur with *are*; similarly, uninflected ‘be’ usually occurs in contexts of underlying *would*-deletion (and not with a distributive/habitual function).\(^2\)

### 6.1.2 CCR in HE and AAE

After it was established that HE speakers have AAE contact features in their repertoires, both the study data (n=665) and a smaller subsection of data collected from AAE speakers in the speech community (n=166) were coded according to the procedures described in sections 3.3 and 4.4 (i.e. in additional to cluster/bimorphemic tokens like ‘picked,’ monomorphemic tokens like ‘test’ were coded for phonological environment and morphemic status).\(^3\) The supplementary AAE data come from sociolinguistic interviews with 7 African American English speakers, aged 12-13, three males and four females, who attended Chewning Middle School. These speakers attended classes with the HE speakers in the study with the exception of an ESL class which met one period a day. All of the AAE speakers reported that they had lived in the Piedmont region of North Carolina (Durham, N.C. or Raleigh N.C.) all of their lives.

A GOLDVARB\(^4\) run on the HE data found a significant effect (p<.05) for following phonological environment, morphemic status, gender and LOR in years. The results for the first three factors are shown in Table 7. The results for LOR are graphed in Table 7.

---

\(^2\) Consult Wolfram (1974), Fasold et. al. (1987), and Bailey (2001) for further reading on black and white speech differences in the South.

\(^3\) Only clusters that ended in a stop and shared voicing were coded, e.g. [st] but not [nt] and ‘and’ and ‘just’ were excluded (see section 4.4.5.). Derivational tokens (*slept, told*) were not considered since they were originally coded as marked for tense if they showed a vowel change with cluster reduction (i.e. *[slɛp] and *[slɛpt] were both coded as marked but not [slip])

\(^4\) GOLDVARB is an updated version of the VARBRUL program described in section 3.4.1.
Table 7: Factor weights for 3 CCR factor groups

<table>
<thead>
<tr>
<th>Factor</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological Environment:</strong></td>
<td></td>
</tr>
<tr>
<td>_C</td>
<td>.52</td>
</tr>
<tr>
<td>_P</td>
<td>.61</td>
</tr>
<tr>
<td>_V</td>
<td>.33</td>
</tr>
<tr>
<td><strong>Morphemic Status:</strong></td>
<td></td>
</tr>
<tr>
<td>Bimorpheme</td>
<td>.54</td>
</tr>
<tr>
<td>Monomorpheme</td>
<td>.63</td>
</tr>
<tr>
<td><strong>Gender:</strong></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>.55</td>
</tr>
<tr>
<td>F</td>
<td>.42</td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td></td>
</tr>
<tr>
<td>= -359.255</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>= 0.031</td>
</tr>
<tr>
<td><strong>Input</strong></td>
<td>= 0.64</td>
</tr>
</tbody>
</table>

Table 7 demonstrates that increasing LOR inhibits overall reduction, though at the upper LORs (5+ years; n= 407) both phonological environment ($\chi^2$= 17.67; p<.01) and morphemic status ($\chi^2$= 13.360; p<.001) still show significant effects. This pattern may be
compared to the phenomenon described in section 3.5, where speakers in Detroit
different in the overall rate at which the CCR rule was applied, but not in overall
constraints (or constraint orders). Additionally, Table 7 shows that there is an overall
tendency for males to apply the CCR process over females (at a factor weight of .55>.42)
which is notable since there was no gender effect for the grammatical variable of tense
unmarking. This is evidence that the CCR variable may well have a social significance in
the speech community that the grammatical effect of tense unmarking does not. The
strands of this social significance may be more complicated to unravel here, however,
since it is not clear 1) what particular social meaning CCR may have in this community
nor 2) whether gender itself is a primary or secondary social index for this meaning
(masculinity, for example, may well be linked to some third factor—school vs.
community orientation, for example). More qualitative/ethnographic evidence would be
needed to establish a precise social meaning for CCR in Durham HE.

Nevertheless, the linguistic constraints alone have important things to say about
speakers' orientation to the community around them. Figure 24 and Figure 25 show both
VARBRUL/GOLDVARB factor weights and raw percentages for CCR reduction for
Durham HE and AAE as well as eight representative varieties of native and nonnative
Englishes. These varieties include Detroit AAE (DetAA) (Wolfram 1969), Mexican
American adolescent English in Gary, Indiana (GME) (Hartford 1975), Philadelphia AAE
and European American English (PhlAA, PhlW) (Labov 2003), Los Angeles Chicano
English (LA CE) (Santa Anna 1991), East Los Angeles (Wald 1981) and Chinese English
interlanguage (ChnE) (Bayley 1991), San Antonio Tejano English (Bayley 1994) and Austin (TX) Chicano English (Galindo 1987). The samples for Philadelphia come from the Urban Minorities Reading Project (UMRP) and include separate factor weights for Philadelphia Latinos who learned to read in Spanish (PhlLat_S) and English (PhlLat_).

Figure 24: Factor weights for CCR by phonological environment for Durham HE and nine English varieties

---

5 Where following C was articulated into subgroups (e.g. obstruent, liquid, glide, etc.), the figures for obstruent were used in the graphs. This convention was followed for data from Bayley (1991, 1994) and Labov (2003).
There are several surprises in these findings. First, while showing the overall constraint ranking C>V, HE actually shows slightly higher rates for pre-pausal tokens (61%/75%), which usually pattern between _C and _V. The effect of pause, however, does show variation even in native speakers: for most of the native dialects of English analyzed by Guy (1980), a following P patterned with a following V; however for a few exceptions, it behaved like a following C. In terms of other Hispanicized English varieties, Hartford (1975), in a study of adolescent Mexican American speakers in Gary, Indiana, reports relatively high rates of CCR for prepausal tokens (70.1%) vs. prevocalic tokens (51.7%), with preconsonantal tokens still (narrowly) showing the highest rates of deletion (72.1%). Though Hartford coded only monomorphemic tokens in order to control for morphemic effect, it is interesting that another U.S. Hispanicized English

Figure 25: Percent CCR by phonological environment for Durham HE and six English varieties

There are several surprises in these findings. First, while showing the overall constraint ranking C>V, HE actually shows slightly higher rates for pre-pausal tokens (61%/75%), which usually pattern between _C and _V. The effect of pause, however, does show variation even in native speakers: for most of the native dialects of English analyzed by Guy (1980), a following P patterned with a following V; however for a few exceptions, it behaved like a following C. In terms of other Hispanicized English varieties, Hartford (1975), in a study of adolescent Mexican American speakers in Gary, Indiana, reports relatively high rates of CCR for prepausal tokens (70.1%) vs. prevocalic tokens (51.7%), with preconsonantal tokens still (narrowly) showing the highest rates of deletion (72.1%). Though Hartford coded only monomorphemic tokens in order to control for morphemic effect, it is interesting that another U.S. Hispanicized English
speech community shows CCR results where _C patterns with _P (with _V showing distinct effects).

Perhaps the more surprising finding is the similarities between the Durham HE constraints and the Durham AA constraints, which mirror each other (factor weights .56/.62/.29 for HE and .59/.46/.37) in terms of the novel rank orders P>C>V. This same effect is found for Bayley’s (1991) study of Chinese interlanguage, where, in terms of raw percentages, P>C>V (clusters are slightly less likely to be deleted before pauses than before vowels). Thus, there is limited evidence (the AAE sample used here only consisted of 7 speakers and 155 tokens) that HE speakers in Durham do show convergence with native vernaculars in terms of the phonological environment constraint on CCR. Furthermore, this result is surprising since both HE and native varieties of AAE show the same (idiosyncratic) rank ordering of constraints.

The results for grammatical category (shown in Figure 26 and Figure 27) provide perhaps even more compelling evidence of HE speakers’ orientation to the groups around them.
Figure 26: Factor weights by morphemic status for CCR for Durham HE and six English varieties

Figure 27: Percent CCR by morphemic status for Durham HE and five English varieties
While the phonological constraints shown in Figure 24 and Figure 25 can be attributed to universal effects of articulatory factors, speakers must ostensibly be attuned to finer, English-specific details around them in order to acquire the morphosyntactic/grammatical constraint. Though the Durham HE speakers do not show the great overall differences in the contrasting rates of CCR for monomorphemes vs. bimorphemes that speakers of native English vernaculars show (including English-dominant LA Chicano speakers), the results in Figure 26 demonstrate that HE speakers are showing the same speech norms as their (AAE-speaking) peers for this variable. Interestingly, Durham speakers contrast with Labov’s (2003, 2010) Latino speakers in Philadelphia: neither UMRP group of 8-11 year old children (n=397), those who learned to read in English nor those who learned to read in Spanish, showed significant effects in terms of CCR for grammatical status. Furthermore, the Spanish group did not show a significant effect for the phonological constraint. In this context, Durham HE speakers show evidence of accommodation to local dialect traits whereas other interlanguage varieties, including Vietnamese English and Chinese English, do not.

These results are further supported by the evidence presented in 5.5 on tense unmarking, which demonstrates Durham speakers are experiencing a steady retreat of the unmarking variable (see Figure 22), for example, 61% for LOR 1 vs. 31% for LOR 15. The additive effect, where learners are more likely to reduce an inflectional vs. lexical cluster due to an additive effect of the grammatical and phonological processes (variable unmarking + variable CCR) does not operate in the same way in this community as
both Chinese English interlanguage and Vietnamese English interlanguage (parallel lines in Figure 27). In this way, the Durham HE community demonstrates it is sociolinguistically distinct from English language learner communities of speakers from both Chinese and Vietnamese backgrounds, as well as showing accommodation to local vernaculars in ways that other Latino communities do not.

6.1.2.1 **Collinearity in morphemic status and phonological environment**

After presenting the results for CCR for Durham HE and Durham HE using raw percentages and GOLDVARB figures, a logistic regression was used to analyze the CCR data in terms of factors morphological status, phonological environment, LOR, and gender—the factors GOLDVARB selected as significant). Surprisingly, the regression results showed a nonsignificant result was returned for morphemic status and gender, leaving only two factors to correlate significantly with CCR variation: phonological environment (p<.000) and LOR (p<.05). Further investigation revealed that an interaction effect contributed to the non-significant result: in fact, phonological environment and morphemic status are themselves correlated, both in Durham HE and in Durham AAE. In both the AAVE and HE CCR data, monomorphemes are more frequent than bimorphemes; in the same way, preconsonantal tokens themselves are more frequent than prepausal tokens, which are more frequent than prevocalic tokens (Figure 28):
We have observed the central effect of frequency on the unmarked tense variable in earlier sections, for lexical aspect and verb type/phonetic saliency (5.3.3-5.3.5), with an interaction for verb class/salience and frequency as well as stative type tokens and frequency. Furthermore, we established a frequency constraint for background/copula tokens in section 5.4. Figure 28 is a striking result since both CCR patterns themselves and the parallel results for CCR in both communities can be accounted for by token type
(mono > bi; _C > _P > _V). The use of inference trees (Tagliamonte & Baayen 2012) establishes that the most important factor for CCR in HE, with all factors included (phonological environment, aspect, morphemic status, school, gender, age, LOR) is phonological context, such that vowels pattern separately from consonants and pauses. For preconsonantal and prepausal tokens, older speakers are less likely to use marked forms; for prevocalic tokens, speakers with longer LORs are more likely to use marked forms (Figure 29).

Figure 29: Inference Tree for Durham HE CCR

These results call into question the older CCR studies (sections 3.3-3.4) which assumed the structural independence of morphemic status and phonological environment and emphasize the importance of considering collinearity among factors when analyzing CCR.
7. Conclusions and Applications

7.1 Review of general findings

This study has demonstrated that variability in both one grammatical variable (unmarked past tense) and one phonological variable (CCR) in emerging Hispanic English in Durham, N.C. is highly systematic and constrained at multiple linguistic levels. Statistical analyses, including logistic regression, have demonstrated that variation in past tense marking is constrained by internal/linguistic, discourse-based, and social constraints. In terms of internal constraints, Durham HE unmarking shows the same kinds of effects of for phonetic saliency demonstrated by previous studies, namely Wolfram’s (1985) study of Vietnamese English, Bayley’s (1991) studies of Chinese English. In addition, the effect of lexical (inherent) aspect significantly affects marking patterns: telic accomplishments and achievements are marked at higher rates than atelic achievements and non-copula states. However, the internal constraints influence not only marking patterns but each other as well: while frequency significantly promotes marking (e.g. super high frequency tokens like copula are almost categorically marked, even by speakers at low LOR levels), high frequency is simultaneously characterized by high phonetic salience. Moreover, in discourse environments where past tense marking is highly favored (e.g. narrative backgrounds), high frequency copula tokens are used at significant rates. In all, learners’ marking patterns show an interdependency of effects characterized by redundancy at multiple linguistic levels (grammar, phonetics, usage) as speakers build proficiency in English.
At the same time that they are developing tense marking patterns, speakers show evidence of acquiring English-specific speech norms which are present in their community. In terms of the phonological variable (CCR), speakers show accommodation to both universal articulatory processes (more reduction before consonants) as well as finer-grained, English-specific patterns (more reduction of monomorphemic tokens). In this way, Durham HE speakers show convergence with native vernaculars not demonstrated by other English interlanguage varieties (Chinese/Vietnamese English), nor other Latino varieties of English (Philadelphia Latino English).

In all, the Durham speakers provide striking evidence for how ethnolect grammars develop by sourcing both idiosyncratic norms present in English interlanguage as well as the local grammars of a speech community of peers. They demonstrate that second language learning and second dialect configuration cannot be thought of as two separate processes. As an ethnolect stabilizes, its speakers incorporate both nonnative (aspectual marking) and native English rules (marking by phonological environment), and these rules continue to operate even for speakers who know very little Spanish. In this way, the Durham speakers provided vivid empirical evidence of language change in terms of language learning, language contact, and ethnolect emergence as speakers configure their own norms while simultaneously incorporating local patterns.

### 7.2 Institutional practice: Identifying English Language Learners

...The transition into the Syntactic Stage occurs, and the child is now able to communicate more effectively with other speakers of the language. The child
continues to expand his or her knowledge of vocabulary, and greater use of adjectives, adverbs, and prepositions, and verbal expressions is demonstrated. Admittedly, this use of language is primitive in nature, but further interaction with speakers of the language provides opportunities to experiment and refine personal skills. (Ballard, Dalton and Tighe 1995: 2)

Language differences as we see them are symbolic of cultural distance. They become instruments of educational failure when they are interpreted in a way to predict and insure this failure. Spears (1987), qtd. in Labov (1987)

The first excerpt above comes from section 1.1 of Examiner’s Manual of the IPT I: IDEA Oral Language Proficiency Test of English, “Theoretical Considerations in Language Acquisition and Language Learning,” the standardized test instrument I used as a K-12 ESL teacher in rural Granville County (N.C.) Schools in 2003-2005.1 With a caseload of almost 100 students, teachers in the county schools were required to evaluate both new enrollees and students already receiving ESL services twice a year in reading, writing and oral skills.

As a linguist, one of the first experiences I had as a new teacher was realizing that an appreciable number of monolingual speakers of local and regional dialects would be scored by the IPT Oral test as English language learners. Fought (2003) describes the same experience in the California high school where she was conducting ethnographic field work. Features like negative concord were evaluated as transfer features from Spanish (‘I didn’t do nothing’ vs. Sp. No hice nada) for students who were monolingual in English—these patterns occurred not due to the influence of Spanish but because of the influence of nonstandard (L1) English.

---

1 As of 2008, the NC Department of Public Instruction had transitioned to using the WIDA Access Placement Test (W-APT) (Yankosky et. al 2012).
I was curious how the HE speakers in the study, especially those who were English dominant or English monolinguals would fare on the IPT Oral test I had used. In order to address this question, I compiled skill items from the syntax strand of the IPT Oral Examiner’s Manual (1995) and then matched these items with examples from the HE corpus, using both my own transcribed data as well as data transcribed by other graduate student fieldworkers in SLAAP (see section 4.3 and Kendall (2007)). Table 8 shows the skill items, taken from Forms A-B of this edition of the IPT-I oral proficiency test (1995: 15, 23 (forms C-D duplicate the same skill items) as well as matched examples from the HE data.
### Table 8: Oral IPT Test by skill area (syntax) with HE examples

<table>
<thead>
<tr>
<th>Skill Area</th>
<th>Speaker/Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Yes/No Response</td>
<td>Tony (12y, LOR 5): She told me ‘0 You tryin to learn English?’ (dps1050d)</td>
</tr>
<tr>
<td>11. Regular Plurals</td>
<td>Cleo (13y, LOR 10 years): I have two sister 0, and that’s all. (dps0050d)</td>
</tr>
<tr>
<td>12. Verb ‘to be’</td>
<td>Paco (15y, LOR 6): His name 0 Mr. Capone (dps0030d)</td>
</tr>
<tr>
<td>17. Pronouns: Subject</td>
<td>Jackson 10y, LOR 8: They had a birthday party. (dps052)</td>
</tr>
<tr>
<td>18. Verbs: Present Progressive</td>
<td>Lori (12y, LOR 10): When we 0 playin’ around, somethin’ like that, we talk in English. (dps0040d)</td>
</tr>
<tr>
<td>19. Mass nouns</td>
<td>-</td>
</tr>
<tr>
<td>20. Conjunctions</td>
<td>-</td>
</tr>
<tr>
<td>21. Negative Statements</td>
<td>Graciela (13y, LOR 7): …and when I woke up I didn’t know nothing (dps0800d)</td>
</tr>
<tr>
<td>33. Verbs: Habitual Present</td>
<td>Marcos (13y, LOR 10): They 0 be like ‘You go first,’ and all that. (dps0110)</td>
</tr>
<tr>
<td>35. Comparatives</td>
<td>Jorge (13y, LOR 13): I’m more loved than them (dps 0402d)</td>
</tr>
<tr>
<td>40. Questions: Future Tense</td>
<td>-</td>
</tr>
<tr>
<td>45. Irregular Plurals</td>
<td>-</td>
</tr>
<tr>
<td>46. Pronouns: possessive</td>
<td>Luisa (9y, LOR 3) [from section 6.1.1]: We made this friendship bracelet, but I forgot mines (dps0280d).</td>
</tr>
<tr>
<td>53. Superlatives</td>
<td>-</td>
</tr>
<tr>
<td>54. Verbs: Irregular Past Tense</td>
<td>Fely (8y, LOR 4): Then Fiona think that Shrek was in her room…(dps1271)</td>
</tr>
</tbody>
</table>

Using only the sentences transcribed for unmarked tense/CCR as well as the few Durham HE interviews partially transcribed in SLAAP, I was able to fill in 10 of the 15 skill areas (67%) with nonstandard English examples which would be scored as errors by the IPT Oral test. To give some context, a second-grader who made nine or more errors in the context of both the syntax and other sections (verbal expression, e.g.
retelling the main ideas of a story or vocabulary, naming animals on the farm) would be designated Non-English Speaking [NES] according to the test. A 6th grader who made four or more errors on this in conjunction with any other section of the test would be designated Limited English Speaking [LES].

This exercise illustrates the importance of linguistically-informed accounts of second language acquisition in multidialectal speech communities. In order to provide effective ESL services to students who have a genuine need, nonstandard English speakers first must be distinguished from English language learners. The first step in this process is a contextualized, accurate, and linguistically-informed account of the language speakers actually have.

Misidentification may not only have negative sociopsychological effects on students, but also far-reaching economic and institutional consequences. For example, North Carolina statewide funding for LEP (Limited English Proficient) students for the 2012-2013 budget was $84,463,502 (Matteson 2013). Thus, if even a 3% improvement were made in accurately identifying ELLs, the savings to the state system could be in the millions of dollars. Alternately, by maintaining current funding levels and effecting a 3% reduction in caseloads, ESL teacher to student ratios could be lowered, allowing for more contact hours of (higher quality) instruction and, ostensibly, sooner exit from the program for all students.
7.3 Future directions

Future studies should consider the effects of additional social variables for HE in order to explore ‘horizontal’ variation in (Preston 1989) (section 2) in interlanguage, including register (the HE corpus contains additional protocols including reading a picture book aloud and a conventionalized translation task), regional variation (Wolfram, Carter and Moriello 2004, Kohn and Franz 2009), and even individual characteristics like speaker agency (Carter 2007) at the individual level. Secondly, while there is convincing evidence that HE speakers both show consistent features of AAE in their speech as well the constraint orders that govern them, the next step in examining accommodation patterns is to compare the comprehensive functioning of other features (e.g. copula absence, negation patterns) in order to understand finer-grained and co-occurring patterns of convergence and/or divergence. Applied research should document how these unitary (nonstandard) English norms which exist in the local speech community can be incorporated into standardized tests and institutional procedures designed to classify students as English language learners (vs. Standard English learners).

Finally, future, cross-linguistic studies should continue to include verb class in their analyses, as it is strongly supported here as a central independent influence on variation in past tense marking. However, researchers should continue to closely examine interactions between factors in their models, especially with regard to frequency, which, in the context of fixed linguistic facts, articulates itself in surprising ways (Erker and Guy 2012). For example, while this study has established the strong interdependence of frequency and phonetic saliency/verb class, future studies might
examine in detail the intersection of phonological environment, morphemic status, and CCR in terms of frequency. While the traditional claim has been that (native) English speakers are preserving grammatical information in reducing monomorphemes more often than bimorphemes, future studies should consider the alternate explanations given in Bybee (2002) and supported here. If -t/-d suffix in regular past tense verbs is frequently (35%-40%) followed by a vowel-initial preposition or particle, as in formulaic sequences looked at, picked up, then a ‘chunking’ effect above the level of morphology may be responsible for very low rates of prevocalic deletion, not morphemic status itself. Given that lexical aspect is coded at the level of the entire VP (‘get on my nerves’ is atelic while ‘get home’ is telic), it would be interesting to look at the influence of aspect vs. V- vs. C-initial particles in verb phrase configurations in the context of acquisition.

While these types of theoretical questions are valuable in and of themselves, we can make the case for sociolinguistic accommodation (in terms of significantly correlated constraints and constraint orders across varieties) without having definitive answers to them. Here, both the quantitative tools and social/critical perspectives of variationists have a vital role to play in future trajectories of SLA research. Our perspectives on both native variation and SLA itself are uniquely informed by noticing what learners (and emerging dialect speakers) notice, exploit, retain, repurpose, and reflect in the context of limited repertoires paired with limitless communicative needs. This study has shown that while listening to others they ultimately make their own rules.
References


Carter, P. 2007. Phonetic variation and speaker agency: Mexicana identity in a North


Biography

Erin Callahan-Price received her B.A. in Linguistics at Yale University, where she did fieldwork in the French West Indies on Guadeloupe Creole, and her M.A. in English (sociolinguistics) from N.C. State. She grew up in Charlotte (N.C.) and is a proud, native North Carolinian. She has taught high school Spanish, French, and ESL in the public school systems in Durham and Granville Counties, as well as English at a bilingual school in Querétaro, México. She is a former rugby player and current beagle-lover.