The Role of Socio-indexical Information in Regional Accent Perception by Five to Seven Year Old Children

by

Erica Lynn Beck

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2014

Doctoral Committee:

Assistant Professor Carmel O'Shannessy, Chair
Professor Marlyse Baptista
Professor Susan Gelman
Associate Professor Robin Queen

ACKNOWLEDGEMENTS

It's hard to know where to start - there's such a long list of people without whom this thing would have never been finished. Thank you to my committee members, who have been very generous with their time and advice. I really appreciate the fact that although I was out of sight the last year, I wasn't out of mind at all. Thank you for being so easy to work with and so very supportive. A special thank-you to Carmel, who read many drafts, and always liberally uses the word "brilliant." It may be dialect variation, but it certainly is an ego-booster!

Thank you to Dr. Samuels, Superintendent of Schools, the Kindergarten teachers and principals at the sites where this study was conducted for allowing me to conduct this study in your schools. I was touched by the large number of families who chose to participate, and how helpful they were. It was an honor to work with you, and I am very proud that I was able to return to my alma mater to finish my formal education where it started!

Thank you to my sister-in-law, Jessica Gray, and her friends who kindly agreed to spend an afternoon recording lists of words for the stimuli. For the rest of the stimuli, I owe a debt to my friends Josh and Beth Lindemann and Christopher and Michelle Boyd, as well as my brother and sister-in-law, Brandon and Kim Hall, who donated their voices to this project.

I am fortunate to have so many wonderful friends, and to have had such a supportive group of people around me there in Ann Arbor. This has been one of the nicest periods in our lives, and you have all played no small part in that. In no particular order, thank you, thank you, thank you to Jen and Neil Warner, Derek and Mara-Bell Mancini-Lander, Steffi and Kay Wille, Eleni and John Van den Bergh, Ginger and Avi Derrow, Annette Bryson and Vittorio Trionfi, Jen and Linh Ngyuen, Reina Kuroshima, Leslie Elliot, Sara Tang, Nikki Williams, Cathy Wood, Mollie Bogan, Judy Powers, Nancy Hagendorn-Fiedler and Eddie Stephens. Seriously, this would have been over years ago without you. I can't tell you how much I appreciate your friendship.

To my Linguistic Department friends who have made this such an enjoyable journey, and continue to help me out even now: Anna Babel, Kevin McGowan, Susan Lin, Jon Yip, Tridha Chatterjee, Harim Kwon, Stephen Tyndall, Tim Chou, Eric Brown and Joseph Tyler. To my more distant friends, Heather, Tiffany and Nancy, who've always been right there, despite never actually being right there, but who were always so outspoken in their support of me. And, perhaps in a dissertation acknowledgement first, thanks to the organizations who have provided wonderful parks to ponder one's dissertation problems: The University of Michigan Arboretum and Botanical Gardens, Pinckney State Park and the Mount Vernon Regional Park. Not all who wander are lost….they may just be (avoiding) writing a dissertation.

Thanks to Dennis Nelson, who never thought it was a bad idea that I leave work to go work on my dissertation. Finishing this would not have been possible otherwise. I am happy to say I now can vacuum the office too.

To my parents, Darryl and Maggie Hall, but especially my mother, who not only regularly drove 500 miles for the privilege of babysitting, but spent hours listening to me complain, and who has formatted and re-formatted this document numerous times now. I admit that I am still dependent on you for many things, the least of which is table formatting.

To my dear daughters, who not only tolerated only getting half of my attention for a while now, but who also piloted all of my experiments, helped to sort my data and volunteered to read and edit the finished product. Isabel and Anja, I am happy to say I can finally answer your questions:

Isabel - No, you don't have to do a PhD, but it was more fun than you think.

Anja - Yes, Mama's book IS finally long enough.

And finally, to my husband James, who encouraged this project from the very beginning, and who has continued to think it was a good idea throughout the process. Aside from being my cheerleader (Sorry! A mascot!), you also did a lot of macro-writing and statistical "proofreading" over the course of it, which saved me a lot of headaches. I'm sure it won't be the last time. I am looking forward to wherever this crazy idea takes us next. I love you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

Introduction

1.1. General Introduction

Regional accents have been shown to evoke social evaluations and expectations of speakers, and social evaluations of speakers have been shown to influence the perception of speech. Explaining the set of associations between social information and linguistic variation has been one focus of recent research, particularly using Exemplar Theory (Johnson, 1997, 2006; Pierrehumbert, 2000, 2003) as a framework for contextualizing the interactions of these factors. However, within this line of research, little work has been done on documenting how children may acquire the social and linguistic categories in the first place, much less what influence social knowledge has on speech perception. This dissertation begins to address those lacunae by investigating what knowledge children have of regional phonological variation in U.S. English, and whether children aged five to seven understand speech style provides social information about the speaker.

Regional linguistic variation is a good kind of variation to examine in a developmental study for a number of reasons. Gender, age and in some contexts, ethnicity, are frequently encountered and also linked to differences in appearance or other obvious markers of that category (Foulkes & Docherty, 2006). Regional linguistic variation in the United States is not correlated with differences in appearance, and typically would be less frequently encountered by children than gender or age-based variation. Even in cases where children have interactions with an individual having a different regional accent, it has been shown that they are aware that the difference is attributable to the speaker's place of residence or birth.

Furthermore, there is some debate in experimental literature about whether regional accents are acoustically perceptible by children between the ages of approximately one year and six years of age (Butler, Floccia, Goslin, & Panneton, 2011; Floccia, Butler, Girard, & Goslin,

2009; Girard, Floccia, & Goslin, 2008; Phan & Houston, 2008). This gives reason to believe that large variation in performance on discrimination tasks could be expected, giving opportunities to examine what factors may contribute to success in discriminating between regional accents, and what roles knowledge of and experience with accent play in discrimination. For these reasons, regional variation was chosen as the subject of this study.

Predictions arising from Exemplar Theory and its models are used as a basis for many (although not all) of the questions asked in this dissertation: what role does experience with regional variation play in discriminating between regional accents? Is ability to recognize a regional accent a prerequisite to discriminating between them? Or can children discover patterns of regional variation even with no prior knowledge about the existence of regional variation? I specifically investigate what kinds of experience and interaction with regional accents are most likely to positively influence the ability to recognize and discriminate between regional varieties. I also ask whether experience with other kinds of variation and accents transfers in discrimination of regional variation. Finally, I also look at whether overt, declarative knowledge about regional variation assists in discrimination in cases where children have no direct experience with a particular regional accent. Answering these questions should provide a clearer picture of how social knowledge supports developing an adult-like understanding of regional accents.

Given the broad range of topics that provide the backdrop for this study, I provide reviews of a range of topics relating to regional accents. First, the socio-indexical knowledge that adults have of regional accents is discussed, followed by an outline of how Exemplar models are currently invoked to account for the influence socio-indexical knowledge has on speech perception. I then look at the predictions this model makes for how acquisition of socio-indexical and linguistic categories might proceed in children. Because there have not to my knowledge been any prior studies testing these predictions, I review general studies of children's perception and production of regional accents, and in particular, I review evidence of how development of social and linguistic knowledge might be interdependent. Finally, I present the research questions that frame the experiments conducted in this dissertation, and give an outline of the hypotheses, in response to those questions.

1.2 Adult Social Knowledge of Regional Accents

Adult speakers of American English have been shown to have a rich knowledge of social categories that are indexed by linguistic variables (phonetic/phonological, grammatical and lexical) in speech, including race, gender and region of origin (Campbell-Kibbler, 2007; Clopper, 2010; Purnell, Idsardi, & Baugh, 1999; Staum-Casasanto, 2009; Strand, 1999). As with many social categories marked in speech, geographic origin is often a proxy for other character traits in individuals, such as the supposed friendliness, likability, status and intelligence of individuals (H Giles, 1970, 1971; H Giles, Harrison, Creber, Smith, & Freeman, 1983; Lambert, 1967; Preston, 1993). The association of a linguistic element in speech with a particular social group are referred to as social indices, or socio-indexical information, in speech.

American English speaking adults are not particularly accurate at perceiving linguistic markers of regional variation when listening to an unfamiliar accent. Studies show that most American adults are accurate at identifying only three major accent regions in the U.S.: Northern, Southern and Western, (Clopper, Levi, & Pisoni, 2006; Clopper & Pisoni, 2004a, 2004b; Labov, 1998; Preston, 1993), and they often fail to hear differences within these three regions. (Note: from here forward, "listeners" and "speakers" will refer to speakers of American English as a first language, unless otherwise noted). Only listeners who had experience living in multiple accent regions of the U.S. could make more nuanced distinctions between the accents of the regions in which they had lived, providing evidence that experience hearing an accent makes a listener more attuned to its phonetic details (Clopper and Pisoni, 2004).

The location of perceptual boundaries for regional accent is highly subjective, depending on the listener's area of residence. When asked to indicate the major U.S. accent regions, speakers from different areas of the United States have drastically different notions about where those boundaries fall (Preston, 1986). For example, speakers from southern Indiana draw the border between Northern and Southern accent regions further south than do speakers from Michigan (Preston, 1986, 1993; Preston & Niedzielski, 1999), and the precepts that cause listeners to categorize a speaker as hailing from any given dialect area depend on the origin of the listener (Clopper et al., 2006; Clopper & Pisoni, 2007; Rakerd & Plichta, 2003) and their experience hearing different regional accents and living in different regions of the country (Clopper & Pisoni, 2004a).

Speakers often define others' accents in relation to how they *think* they themselves speak, and how "correct" or "standard" they think their native variety is (Fridland, 2010; Niedzielski, 2010; Preston & Niedzielski, 1999), regardless of the acoustic characteristics of their speech. In other words, perception of an accent is subject to change depending on the listener, context, ideologies and non-linguistic information about the speaker. In fact, the presence of non-linguistic social information about a speaker can trigger the perception of an accent, even when it is not present (Hay & Drager, 2010).

Experimental studies provide evidence of listeners' perception of accents being influenced by non-linguistic information about the speaker. It appears that listeners are easily deceived into hearing phonetic features not in the speech signal by manipulating their ideologies about regional accents. For example, Niedzielski (1999) showed that U.S. listeners would report hearing a particular vowel quality characteristic of Canadian accents, even when that vowel quality was not present, simply by being told that the speaker was Canadian. When listening to the exact same recording, if listeners were told the speaker was from the U.S., they would report not hearing the stereotypically Canadian vowel in question. Similarly, listeners also will shift the boundaries of their vowel categories based on exposure to pictures, symbols or other representations of a group of speakers (Hay & Drager, 2010; Hay, Warren, & Drager, 2006). Perhaps the most egregious example of perception being influenced by social expectations is a study that finds listeners will report hearing an accent from a speaker they previously described as unaccented when a picture of the purported speaker leads them to believe that person is of foreign origin (Rubin, 1992).

1.3 Exemplar Theory and Adult Perception of Regional Accent

The phenomenon of speech perception being influenced by socio-indexical information about speakers has been explained via Exemplar Theory (Johnson, 1997, 2006). This theory posits that the brain stores memories - called traces - of having heard a specific token, be it a phoneme, grammatical particle, lexical item etc. These tokens are subsequently abstracted to form an exemplar of particular phonemes, lexical items, etc. (Johnson, 1997). The phonetic details retained would help not only overcome the problem of individual variation in speech perception, but because traces are stored with information about the speaker and social context,

they would serve to help identify individuals and social groups based on the phonetic details present in their speech.

Parallel to the linguistic categories are social categories, which contain detailed information about how their members speak. The most important function of these parallel social categories in speech perception is that they retain detailed phonetic information about productions beyond what would be needed to identify the lexical item, phoneme, etc. This information is eventually abstracted away from traces of interactions with specific individuals and stored in a more general template for speech typical of that category (Johnson, 2006; Munson, 2010; Pierrehumbert, 2006; Squires, 2011).

Having parallel social and linguistic categories receiving activation during speech perception causes them to exert influence on one another. When a listener knows that the speaker belongs to a particular social category, the expectation of how that person speaks will bias what the listener reports hearing. The process also works in reverse: several studies have shown that hearing acoustic markers associated with specific social groups leads listeners to assume the speaker belongs to that group, or at least possesses stereotypical qualities of the group (Campbell-Kibbler, 2007; Hay, Warren, et al., 2006).

Information about how members of a particular group stereotypically speak also has been shown to help listeners to disambiguate input. For example, Staum-Casasanto (2009) finds that listeners will disambiguate words that are homophonous in two ethnolects of American English based on their knowledge about the race of the speaker. For example, *mass* and *mast* are homophones in varieties of African-American English, whereas in standard Caucasian American English, they are distinguished by the pronunciation of the final /t/ in *mast*. If listeners believed the speaker was African-American, they expected /mast/ and /mass/ to be homophones, but not in the case where the speaker was believed to Caucasian. If the information about how people of different races speak were not available to listeners, they would have no reason to preference one interpretation of the token over the other. Similarly, Foulkes et al., (2010) showed that adults could use a regional phonetic variable, the use of which patterns differently for males and females, to help identify the gender of pre-pubescent children. This indicates that listeners know which gender is statistically most likely to use those particular regional variables, and use this information to identify the gender of a speaker in the absence of differences in fundamental frequency. What was not explicitly asked in this study was, however, if the listeners making the

identifications had overt expectations about who would use which variable more, and whether these beliefs affected their choices in the task.

Strand (1999) provides further evidence of the influence of social category information on speech perception in a study testing gender typicality of voices. Voices that were rated more typical of the speaker's gender were processed faster than those that were less so. These results were interpreted to show that listeners have expectations for how members of each gender speak, and are most efficient when the input matches that stereotype.

The details of Exemplar Theory have yet to be worked out in full. There is some debate over the linguistic level(s) at which exemplars are stored (Johnson, 1997; Pierrehumbert, 2003). Specifically, there is some debate whether listeners store individual phonemes, words, utterances, or any permutation of the above. In addition, the general assumption seems to be that direct experience hearing a particular kind of variation is needed in order to form representative categories, although the "critical mass" of experience hearing a variety that is needed to form a category remains unknown. It is also unclear how stereotypes are linked to exemplar categories. These last two questions are germane to the current study, since children are learning about their linguistic and social environment, accumulating tokens for which they may have no labels, and getting exposure to stereotypes and evaluations embedded in their culture of which they are not overtly aware.

Before I more carefully address these issues, I should distinguish between variation above and below the level of consciousness (Labov, 2001). There are kinds of variation about which speakers are aware of and comment on, and as a result can consciously manipulate. These are said to be above the level of conscious awareness.

There are also elements of variation that are well attested by sociolinguistic studies, and pattern with age, gender, class etc., but the speakers are not generally aware of, and do not comment on. These are said to be below the level of conscious awareness. Sociolinguistic processing studies cited earlier in this chapter deal with perception of linguistic variables both above (e.g. Niedzielski (1999), Hay, et al (2006) and Hay and Drager (2010), etc.) and below the level of consciousness (e.g. Strand (1999)). In both cases, perception of sociolinguistic variables can be influenced by socio-indexical knowledge. Use of above-the-level-of-consciousness variables in experimental studies appears to be more common.

In contrast to the actual sociolinguistic variable used in the above-mentioned studies, there is a second set of categories that can be above or below the level of consciousness. The social categories that the linguistic variables are linked to in those studies are usually overtly known to the listeners: gender, age, nationality, and race. To my knowledge, in only one study are categories referenced where it is not clear that listeners are overtly aware of the existence of that category, if it exists for them at all. For this dissertation, it is important to ascertain the effects of an unknown social category on perception, since it is possible that region is not a social category for some of my subjects. For this reason, I review in detail the one adult perception study in which the social categories may not have been overtly identifiable by the subjects.

Drager (2010) conducted a perception study on use of the word "like" by high-school-aged girls in New Zealand, after conducting an extensive ethnographic study, in which she found robust patterns of variation in their production of "like" depending on the girls' social group. However, when the same girls were asked to identify which group or individual was most likely to have used a particular variant in a perception test, their responses seemed to be guided by stereotypes of each groups' character. It appeared that they were not sensitive to the fine phonetic details that distinguished the variants and that those variants indexed particular social groups in the school. Drager explained this result by saying that the more canonical variants were associated with the "good" group of girls because they were stereotyped as being more correct or good, whereas the non-standard variants were attributed to the less mainstream group of girls, although this did not accurately reflect the variants' use.

In the Drager study, the subjects don't have strong associations between the two social groups she identified and particular phonetic realizations of "like." Instead subjects are operating on stereotypes of particular individuals and their group of friends. One possible explanation is that if the category itself is not labeled, i.e. exists below the level of conscious awareness in the social awareness of the listener, it is unavailable to be associated with specific kinds of speech.

If the social group itself is not overtly labeled or identified by the listener, it could be that the association of tokens of "like" is at an individual level, and they have not been abstracted to represent an entire social group. This may be substantiated by the fact that many of the listeners in Drager (2010) tried to identify the individual doing the speaking, instead of the social group, as the experimenter had intended.

In the Drager study, the social category being targeted was locally-based; that category only existed in the school, and would not be recognized by the outside community. Therefore, such a social category receives no reinforcement outside the school context. As a result, it may not be as robust a category as more broadly applicable categories, such as gender, age, etc.

It also may be that Drager interpreted the categories differently than the girls do; they may have seen them as loose groupings of individuals, not concrete social groups that constituted a category. In the 2010 paper, Drager notes that the girls referred to themselves as "different" and normal, whereas the task asked them to identify who ate lunch in the Common Room (the "normal" girls did, whereas the "different" girls did not). From this information, it is possible that they associate the different realizations of *like* with individuals, and that "different" and "normal" are individual characteristics, and not concrete social groups. An alternate possibility is that "normal" and "different" are the social-indexical group labels with which *like* is associated, and the connection to the lunch rooms was either not associated with those groups, or that it was a social meaning not consciously available to the subjects in the study. In either case, if the social category of Common Room Girl and Non-Common Room Girl was not available to the subject, *like* could not index it, and the subjects were left to devise another way to provide the information that the experiment sought. Alternately, even if *like* were statistically associated with where an individual ate lunch in exemplar representations, listeners may not have access to that information when overtly asked to make statements about its distribution.

This is very much a hypothesis on my part, and further information would be needed to corroborate this account. However, this is an important question to address in studies with children, for whom linguistic and social categories are developing, and whose interpretations of those categories may differ from an adult experimenter's. In the following chapters, I describe how in the experimental methodology I take into account the social categories children may have constructed with their limited experience.

The other open question about Exemplar Theory asks about the role of stereotypes in perception. There are social groups with which listeners will have no direct, personal experience, but rather only know through characterizations and stereotypes. For example, Southern accents in the United States are generally quite salient to listeners, although the listeners themselves may not have ever interacted with a person from the South. Do listeners

form a category for Southern accents without interacting with a Southerner? Will exposure to stereotypes or media representations of another accent also effect category formation?

In identification tasks like Drager's, there is evidence of stereotypes overriding any actual experience with a particular social variable. Hay, Nolan and Drager (2006) and Hay and Drager (2010) find effects of stereotypes of Australian speech in vowel perception experiments with New Zealand listeners. Niedzielski (1999) found Detroiters showing a perceptual bias when it came to perceiving vowels that were stereotypically Canadian. Foulkes and Docherty point out that statistical frequency does not account for all of perception (2006), but how other kinds of stereotypes and overt knowledge either form or interact with experiential exemplar categories has yet to be thoroughly explained.

These two questions are particularly relevant when considering how children acquire both linguistic and social exemplar categories, given that their experience with both the linguistic and social worlds is limited and still evolving, and thus subject to constant re-evaluation. Below, I review some of the predictions made about how children's acquisition of social and linguistic categories would proceed under Exemplar Theory. I then review the small body of studies that address children's perception of regional accents. Although none were conducted specifically from an Exemplar Theoretic perspective, I look for clues that may address some of the questions that are raised when considering Exemplar Theory from the perspective of acquisition.

1.4 Predictions of Exemplar Theory about Children's Social and Linguistic Categories, vis-a-vis Regional Accents

Munson (2010) suggests a model of acquisition, adapted from Beckman et al., (2007), by which a child collects traces of interactions with known individuals, called encodings, which are linked to specific speakers that the child has heard. The "generalizations" or social characteristics of those speakers (male, African-American, middle-aged, etc.) would begin to populate the set of social categories available to the child. Presumably, the speech of the individuals from whom the social labels were taken also constitutes the basis of the speech exemplar for that category.

The model does not address how children establish categories for social characteristics that they do not yet perceive in their social world. Foulkes (2010) asserts "category labels develop over statistical regularities." The importance of having a label for the category is

underscored in this claim, perhaps indicating that children can only form categories for social types of which they are explicitly aware.

This might mean that they effectively ignore any variation coming from groups of which they are not overtly aware; if no label is available for the social group, then they would not know those phonetic details have any importance in identifying the speaker. From a developmental viewpoint, however, it would be beneficial if children did retain as much phonetic detail as possible, and as soon as the association is made between speech and social type, have the stored traces of speech available to them. This may be accomplished by Munson's "library" of talkers, allowing them to retain variation as individual variation, but reassign it later when they identify a more general social category.

Foulkes and Docherty (2006) posit that children can reinterpret their social worlds during the course of development, re-weighting and re-defining the social categories most relevant to them. Thus, it may not be the sum total of tokens that determines the categories, but its relevant salience and social importance to the individual. This, of course, can change throughout the lifespan, so retaining the details associated with individuals, all of whose social characteristics and labels are not known, may act as a way of bridging between accumulated tokens of variation and formation of a category, even in adults.

Foulkes and Docherty (2006) also note however that for the more arbitrary, less apparent, social classifications, especially those to which the child has little exposure, it is possible the child will never reliably recognize that kind of variation. This seems to suggest that it is only direct experience with a variety, and only experience that reaches some unspecified threshold of salience or importance that will result in construction of a social category.

Although matched guise perceptual dialectology studies have been conducted with children (Day, 1980; H Giles et al., 1983), to my knowledge no studies have examined whether children's knowledge of a speaker's home region biases or affects speech perception. None of the hypotheses or models discussed in the preceding paragraphs address what effect nascent social categories might have on, for example, perception of vowels, so it is unclear if those two sets of categories would have the same mutual influence in children's perception of social variation as they do for adults.

Finally, the question regarding the role of overt knowledge or awareness of variation, even when it has not been directly or intensively experienced, also remains unanswered with

respect to children. Children are consumers of media in which different kinds of social variation are portrayed, and also may be exposed to stereotypes or imitations of other varieties from adults. It is unknown whether children begin to form expectations of how people belonging to different social categories sound based on those representations, regardless of their accuracy or consistency. I attempt to address that question in this dissertation by comparing the effects of both "overt" and "covert" knowledge of regional accents on discrimination.

1.5 What Is Known about Children's Perception of Regional Accents to Date

1.5.1 Acoustic Discrimination

Research on children's acoustic discrimination of regional accents has suggested that they go through a phase of language acquisition in which they are not attentive to regional variation in speech. Between three and five months of age, children are able to distinguish the local accent from another accent in their language (Butler et al., 2011; Egerova, 2010; Kitamura, Panneton, Notley, & Best, 2006; Nazzi, Jusczyk, & Johnson, 2000), but cannot discriminate two unfamiliar accents of their language from one another (Butler et al., 2011). Several studies report that between eight and eleven months of age, the ability to distinguish between local and non-local accents declines (Kitamura et al., 2006; Phan & Houston, 2008), and this "deafness" appears to extend through at least 30 months of age (Phan & Houston, 2008), and possibly until after the sixth year of life (Floccia et al., 2009; Girard et al., 2008). However, Nathan et al. (1998) show that four year olds from Southeastern England have difficulty recognizing common words spoken in a Scots accent, which would seem to suggest that some regional variation is apparent enough to impede comprehension.

Relatively fewer studies exist on accent perception beyond infancy. Two studies, one of native French-speaking and the other of native-British English speaking children aged five and seven, showed children aged five performed at chance in discriminating sentences spoken in their own regional accent from another, but could reliably discriminate between them by seven years of age (Floccia et al., 2009; Girard et al., 2008).

I suspect there are two sources of difficulty for the five year olds in these studies, and the four year olds from the Nathan, Wells et al. (1998) study mentioned above. First, in the Floccia et al. and Girard et al. studies, the fact that children had to rely heavily on their short-term memories in order to complete the tasks made the tasks quite difficult. Unlike in the infant

studies, where it is thought that infants are using only prosodic and rhythmic cues in the sentences to distinguish between the accents (Nazzi et al., 2000), five year olds are also interpreting the sentences for content and segmental information, increasing the amount of processing they were doing while completing the task. These tasks also required children to sort sentences by accent, meaning that they must compare each sentence with all the sentences that they heard previously in the task to try to find similar features on which to group them. Sentences provide a lot of segmental, supra-segmental and semantic information, and having to sort and retain so much detail may have simply been too hard for the five year old subjects.

On the other hand, in the Nathan, Wells et al. (1998) study, hearing single words with no context led to misidentification and failure of word recognition by four year olds. Therefore, simply reducing the amount of speech children hear does not necessarily reduce the difficulty of the task, as the children may not recognize the words without context, and assume that they are hearing novel or nonsense speech.

1.5.2 How Children Interpret Variation and Accents for Social Meaning

Children begin to express preferences for the prestige variety of their language and the speakers of those varieties around age five (Cremona & Bates, 1977; Day, 1980; H Giles et al., 1983; Millar, 2003). However, it seems that children are confused about what features distinguish the prestige variety from the non-prestige variety where one exists (Cremona & Bates, 1977) and sometimes misidentify features of the local dialect as belonging to the standard variety (Millar, 2003). Children between five and eight years of age sometimes expressed negative opinions about the non-prestige variety, although they themselves were speakers of that variety (Millar, 2003).

In another study looking at variation at the lexical (as opposed to phonological) level, Odato (2010) found that children aged seven to 10 knew when the focus marker and quotative particle "like" were being used grammatically. However, they weren't able to associate its use with females, as adults do (Dailey-O'Cain, 2000), meaning that they had not learned about the social patterning of "like" use. Additionally, the children only began to give negative evaluations of "like" usage at age nine to 10, and even then, only the females, suggesting that although they acquired grammatical "like" much earlier, its social significance did not attach until nine to 10 years of age.

The studies cited here deal mainly with grammatical and vocabulary differences in language varieties, yet show that children aged five and six did not have a firm grasp on what distinguished one variety from another, and were unable to associate use of a stigmatized grammatical feature with a specific group. However, children have been shown to produce features of regional (and other kinds) of variation from almost as soon as they can talk (Roberts, 1997a, 1997b; Roberts & Labov, 1995; Shatz & Gelman, 1973). Even more interestingly, not only do they produce that variation, but they do so in a manner appropriate to their age and social cohorts. Below I review those studies, and propose an explanation for the seemingly incongruent findings from the two sets of studies.

1.5.3 Children's Productions of Regional Varieties

Children have been shown to produce phonetic changes specific to their region of residence from the age of three (Roberts, 1997a, 1997b; Roberts & Labov, 1995). Interestingly, they not only mirror the parents' rates of usage of a particular feature, but appear to be advancing the change; that is, using it in ways different from the parents, but in line with their age cohort (Roberts & Labov, 1995). Foulkes et al. (2005) found that children aged two to four in Newcastle-Upon-Tyne matched the frequency with which their mothers used preaspiration in prepausal /t/, and by 3;6 were patterning with their gender in use of this variable (Foulkes, 2010). Similarly, Roberts (1997a) found boys and girls patterning differently, but in an adult-like manner, in their deletion of final /t/ and /d/ as young as three years old.

The above-mentioned studies deal with acquisition of a regional dialect that is the child's native variety. As such, it may be unsurprising to find them acquiring regionally specific features of speech at an early age, since that variety comprises the majority of their input. However, the fact that they are able to not only use the phonological features in the correct contexts, but also adapt their rate of usage based on gender shows sensitivity to patterning of sociolinguistic variation in speech. It also indicates that children may be sensitive to statistical distributions of variables in their environment, although Foulkes, et al. (2005) find that mothers appear to adapt their use of some variables to the age and gender of their children when speaking to them, modeling use of the variable typical of the child's gender in the community. This may indicate that children acquire these patterns not by noticing their distribution in the larger environment, but from imitating the patterns found in child-directed speech.

Several studies have addressed the question of whether all regional phonological variables are acquired equally well by immigrants into a new dialect region. Payne (1976) conducted her dissertation research in a town five miles from where the research for the current study was conducted, documenting the acquisition of the local speech variety by children whose parents moved to the town from outside the Philadelphia area. She found that phonological patterns that were "across the board" in the Philadelphia variety, meaning that they were consistently applied in all phonological environments, were learned by almost all children. The patterns that were conditioned by lexical context, such as the now-famous Philadelphia short /a/, were not fully acquired by any children whose parents were from outside the area. She did find, however, that similarity of the parents' native dialect to the target Philadelphia dialect played a role in the success children had in acquiring some features.

Similarly, Roberts (1997) also found that pre-school aged children were most likely to adopt Philadelphia-specific sound changes when both parents were from Philadelphia, although this was dependent on the sound change in question.

The influence of the parents' native variety has been shown to be much reduced once young children reach school age. In another study of a children moving into a new dialect region, Kerswill and Williams (2000) studied the dialect features present in the speech of children between four and 12 years of age. All of the families had moved to Milton Keynes, England from other areas of the country. Four year olds, who had yet to start school, generally mirrored their parents' regional dialects. By eight years of age, they had lost many features of their home dialects, and more closely mirrored their peers' productions. By 12, no trace of the home dialect was in evidence in production of the variables examined in their study. This is evidence of re-weighting of features in production over the course of development, from the most statistically preponderant to the most socially valuable in that context. It also belies an acute attention to detail- not only phonetic detail, but to the social value of using particular dialect features with their peers. However, to my knowledge no examination was made of whether the children were aware of any of the dialect features that they were acquiring and using, and it is possible all of the variants studied were below the level of conscious awareness.

There is one caveat about the ability to learn and produce a new regional variety: acquisition of some features may also be subject to age of arrival effects. Chambers (1992) found that for six Canadian children who immigrated to England between the ages of nine and

17, the acquisition of the new dialect was uneven, with dialect-specific lexical items being acquired more quickly and evenly than phonological rules of the new dialect, some of which were not acquired at all.

These studies provide evidence that children are at least 'covertly' aware of regionally-specific phonological variables, and how the use of those phonological variables patterns with gender and age. It would therefore be surprising if children were completely insensitive to regional differences in speech, and perhaps suggests that previous findings of the inaudibility of regional differences might be the result of the design of the experiments or stimuli, or that children have a difficulty reporting what they are hearing. These possibilities will be addressed fully in subsequent chapters, and accounted for in the design of the present study.

1.6 Goals of this Dissertation

The overall goal of this dissertation is to gain a detailed understanding of children's awareness of one kind of social variation between the ages of five and seven. Because this is a developmental study, I chose a type of variation that is thought to be emergent in early-school age children, regional variation, and specifically, regional accents. In particular, I examine both the ability to recognize it, as well as to report hearing it and interpret it for social meaning. This allows me to also look at how exposure to variation, regional and otherwise, affects children's ability to discriminate between regional accents. It also allowed me to directly ask the subjects what they know about regional variation and whether this overt awareness of regional variation has any effect on discrimination of regional accents.

The studies above show that children show sensitivity to regional variation in their speech production, but when directly queried may have trouble identifying a regional accent or discriminating between even familiar and unfamiliar regional accents. This distinction is key as Exemplar Theory appears to presuppose an overt recognition of a variety in order to show an effect of social category biasing speech perception. In this study, I try to consistently separate the overt recognition of regional accents from the covert, in order to examine the effects of one on the other separately.

Because it remains an open question as to whether children aged five to seven even discriminate between regional accents of their native language, the first experimental study asks whether children can discriminate acoustically between them using an ABX discrimination task.

Discrimination between speakers with different regional accents is the central experimental methodology used in this study. This was done because success in discrimination indicates that children of this age hear and recognize regional accents as something common across a group of people; that is, they are not attributing any differences they hear to either individual or some other source of variation. In this first task, no reference is made to either accents or regions, in order to find out whether children would identify speakers as having commonalities based on regional accent, even when they are not explicitly prompted to do so.

The second question is whether children use information indexed by regional accent to discriminate between speakers. In contrast to the first task, the second task explicitly references the social meaning of the regional accent by asking children to identify another member of their geographical community based on accent. In this task children hear two speakers, one representing each accent, and must choose which speaker they think sounds most like themselves in each trial. By doing this, I was testing whether referencing the accent affected discrimination, either positively or negatively. Given that in past studies children aged five and six were unable to match speakers by regional accent, I had not ruled out the possibility that when explicitly asked to find similarities among any set of speakers that children would devise their own heuristic for doing so. Thus, comparing the results of these two tasks would allow me to understand first, if regional accent was a kind of variation they recognized, and if they could use it to find similarities between speakers. If not, it would be possible that regional variation was unknown or not identifiable for children of this age.

The results of these two core experimental tasks are only informative about whether children can discriminate between regional accents and report that they heard a difference, and whether they can mobilize socio-indexical knowledge to assist them in discrimination. However, these two tasks are accompanied by a third in which children answer questions explicitly asking them about regional accents, and are asked to identify the region of origin of the speakers they hear in the experimental tasks. The information from this set of questions allows me to interpret the results of the two experimental tasks to see whether declarative knowledge related to regional variation influences their discrimination ability. I foresee several possibilities: either that they have no knowledge of regional variation, and they are simply guessing which speakers match (or have found some other voice quality, by which to match speakers); that they have a knowledge of regional variation, but can't identify any specific features that characterize such variation and

thus cannot perform the discrimination task; or that they are at least able to identify and differentiate local from non-local speakers.

The third task addressing the state of children's overt knowledge about regional accents also allows me to differentiate between the roles of overt and covert knowledge in discrimination. I might find that children can discriminate between regional varieties, but are unable to answer overt questions about regional variation, or unable to interpret the socio-indexical information encoded in the accent. Depending on the results of the experimental and question tasks, I could find interesting distinctions between the two kinds of knowledge, and correlations in the development of both kinds of knowledge. Given that overt, or above the level of consciousness variables seem to play a different role than the below the level of consciousness variables in sociolinguistic perception studies, the relationship between the two is worth exploring further. It may be that overt expectations or stereotypes of speakers only influence overt reporting tasks, whereas in tasks where no report on the accent is made, the experiential information collected in exemplars may not be over-shadowed by overt knowledge.

Finally, I asked parents to complete an extensive survey on their child's language background. The parents of all children participating in this study completed a survey. This allows me to analyze the results to all three tasks on a within-subject basis, looking for correlations between specific kinds of exposure to, and experiences with, regional variation, and results of the three tasks. It also allows me to examine whether experience with other kinds of variation generalizes to assist in discriminating or recognizing regional variation. The survey data will help differentiate the effects of direct experience with a regional accent through interaction, as opposed to exposure through media in discrimination and identification, and whether direct interaction is superior to television or media exposure to accents. Additionally, it allows us to compare the roles of overt awareness and knowledge about regional accents (both specifically and generally) and experience hearing regional accents, and whether one of these two types of knowledge affect discrimination in either task presented in the study.

Generally, the study can be divided into an investigation of effects of overt and covert awareness of regional accents, with two instruments assessing the knowledge children already possess and two tasks testing children's ability to utilize that knowledge to discriminate between accents, as summarized in Table 1.1

Table 1.1 Relating Dissertation Tasks to Overt and Covert Awareness

Overt	Covert
Awareness Task	Parent Survey Data
Similarity Judgment Social Index Discrimination Task	ABX Discrimination Task

1.7 Research Questions and Hypotheses

Below, I present the research questions and the hypotheses I have generated in response to each:

Q1: Are five to seven year old children able to discriminate between a familiar and unfamiliar regional accent in their native language?

Q2: Are children able to use the socio-indexical information encoded in a regional accent to discriminate between speakers?

Q3: Is ability to discriminate between regional accents at all dependent on having direct experience with the regional accents in question?

Q4: Is ability to discriminate affected by children's overt awareness (i.e. ability to identify a regional accent and state what the regional accent indexes) of the accents in question?

The hypotheses responding to each of the above questions are as follows:

H1: Five to seven year olds will be able to discriminate between regional accents acoustically, when presented with a task that does not suggest any association with regions, accents, etc. This task should show that children are aware that regional variation is common across speakers, and that the differences between the stimuli speakers is not the result of some other kind of social variation.

H2: Five to seven year olds will generally be able to identify members of their local community based on accent. The one caveat to this may be children who do not speak the majority regional accent, as the framing of the question asks children about similarity between themselves and the stimuli speakers. This task should show that children are aware of the socio-indexical information encoded in speech, and can find commonalities between speakers based on that information.

H3: Given that Exemplar Theory has emphasized the role on storing and abstracting tokens of speech in the creation of categories, I would expect that children who have direct experience with the non-local accent to be superior in its identification as compared to children without that experience. In particular, children who have

family members from the South will have a higher rate of accuracy recognizing and grouping speakers with Southern accents.

All children in this study will have had intensive exposure to the local regional accent represented in the stimuli. However, I may find differences between children whose parents are not originally from the town where this research was conducted, or children whose parents are speakers of another variety of English (i.e. African-American English, L2 English, etc.) or another language entirely. These children, although they will have heard the local variety extensively in school, will have had most of their exposure prior to starting school in another variety. Therefore, I may find evidence that their exemplars of local speech may be defined differently than those children whose families speak the local variety, and this in turn may influence the correlations found between experience and discrimination of the regional varieties.

H4: Given that there are effects of speaker stereotypes in adult perception studies, I expect that children who have an overt knowledge about accents also use this information when making decisions about who sounds similar. However, I expect relatively few children of this age to have an overt abstract knowledge about regional accents and their distribution, as they may lack the geographical awareness to comprehend relative distances and location, making geographically based variation uninterpretable to them.

I expect to see less influence of this knowledge in the first task, where no reference to the socio-indexical value of the accent is made. However, I would expect in the second task that children who are overtly aware of regional variation to draw on this knowledge to find speakers from their community, since they will be having to access information about the speaker to make this match, instead of (possibly) relying just on acoustic similarities.

I also expect children who are overtly aware of other kinds of variation, be it ethnic or L2 variation, likely don't transfer that knowledge to regional variation, since its sources are different, and regional variation still requires an understanding of geography to interpret.

1.8 Structure of the Dissertation

In the following chapter, I present a demographic and historical sketch of the community in which the research for this project was conducted. I also present the data collected from the Parent Questionnaire, and aggregate it into categories that give descriptive data about the levels of experience children have with regional and other kinds of variation, as well as their residency backgrounds and parents' residency backgrounds.

The third chapter will present the results of the Awareness Task. I review each of the five questions asked in this task, and present the results of each for the entire subject group as well as by sub-groups. These sub-groups are identified based on the same demographic and linguistic differences that represent levels of exposure that members of each group have to the local regional variety in this town. Finally, correlations between responses to the Awareness Task questions and survey data on exposure to regional accents are examined.

Chapter 4 presents the first experimental task, the ABX discrimination task. This task asks whether children can discriminate between regional accents without any prompting or suggestion. The results of the task are examined for correlation between both the Awareness Task questions (overt knowledge) and the Survey data (covert knowledge/experience) to see what role both kinds of information may play in assisting or complicating discrimination.

Chapter 5 presents the results of the second discrimination task, in which children are asked to use socio-indexical information to identify speakers. In this task, they hear two speakers, one with the local accent and one with a non-local accent. They are asked to choose the speaker that sounds most like themselves, and despite mismatches in gender and age, the results should show whether they will use indexical information about the speaker's home region to identify individuals from their community as sounding most similar to themselves. These results are also examined for correlations with survey data and Awareness Task data. Particular attention is paid to demographic data, given that a number of children in this study would have reason to NOT find similarities with the local stimuli speakers, based on their ethnicity, gender and home regional varieties.

Finally, I present a recapitulation of the findings of the tasks and statistical analyses, and discuss their methodological and theoretical implications.

CHAPTER 2

Subject Demographics and Parent Questionnaire

2.1 Subject Population Context

The study was conducted in a school district serving a town which I will call "Northville" the purposes of this dissertation. The town is located six miles northwest of Philadelphia, in Montgomery County, Pennsylvania. Northville is an urban area (as opposed to a suburb or bedroom community outside of Philadelphia), and the latest available demographic figures show that the population numbers 34,324 in the Borough of Northville (U.S. Census, 2010). In the 2010 Census, Northville Borough is reported as being 40.9% Caucasian, 35.9% African-American, and 28.3% Hispanic. Since the last census in 2000, the Hispanic population has almost tripled percentage-wise (from 10% to 28%), with the Caucasian residents decreasing percentage-wise from 54% to 40% of the population. The African-American community has remained relatively stable over the last decade (34.8% in 2000 to 35.9% in 2010). The annual median household income was $35,714 in 2010 (U.S. Census, 2010).

Two neighboring townships, East and West Township, also are served by the same school district. East Township had a reported population of 13,590 in the 2010 Census (U.S. Census, 2010). Here, 82% of the population is Caucasian, 9% African-American, and only 3.2% Hispanic. These proportions have remained constant over the past 10 years. Median household income was reported as approximately $60,000/year.

West Township has a population of 15,633, and according to the last available statistics, 89.5% of those residents are Caucasian, 6.1% African-American and 1.6% Hispanic (U.S. Census, 2010). The median household income was $63,613 in 2010. By comparison, the Montgomery County as a whole has a median income of $78,446 and is 82.4% Caucasian, 9% African-American, 6.6% Asian and 4.4% Hispanic (U.S. Census, 2010). Northville and its

surrounding townships are considerably less well off financially and are more ethnically diverse than the rest of the county.

The Northville Area School District serves 6,800 students between grades K-12 from Northville Borough and East and West Townships. There are six elementary schools in the District, each with its own unique demographic profile. The school district busses its students, meaning that students are transported outside their immediate neighborhood to schools in other neighborhoods. This is done to increase the ethnic and socio-economic diversity in all of the schools in the district, such that no school is more affluent or less diverse than the other schools in the district. As a result, both elementary schools participating in the study serve students from the Northville Borough as well as East and West Townships.

2.2 Parent Language Questionnaire

As part of the study, I sent an extensive questionnaire about demographic data and past exposure to different languages and accents to parents of prospective subjects, along with the IRB Informed Consent paperwork. The parents were asked to fill out and return the questionnaire, and were given a $10 gift card as a thank-you.

This questionnaire was based on a genetic history questionnaire used in M. Baptista and P. Verdu's project: "Reconstructing the ancestry of Cape Verde founding populations," and modified to determine the amount of exposure children had to different languages and accents in the home environment. The complete questionnaire is reprinted in Appendix 1.

2.3 Description of the Questionnaire Items and Design

Based on my hypothesis that both exposure and meta-linguistic awareness of regional accents would enhance children's ability to report hearing them, I tried to ascertain what possible factors might contribute to both exposure to and awareness of regional accents in a five- and six-year-old child's life.

For the regional accent exposure questions, I asked parents to estimate specific amounts of time children had spent traveling, hearing regional accents in the media or interacting with speakers from other regions. I ask for information about both the child and the parents: where they were born, raised, and currently reside. I also ask about any friends or family members who speak with different accents (both regional and foreign), and the amount of exposure the child

has to those individuals. Parents were asked to provide specifics as to who the speaker was, how often the child saw him or her, and whether that person had a close relationship with the child. I realize that this is potentially a flawed question: the parents may not consider a speaker to have a different accent, although I might, and vice versa. The result could therefore be underreporting of children's exposure to other accents. However, short of conducting intensive observations of the children during their daily routines, I did not see an alternative way to assess this kind of exposure. I also attempt to assess the level of closeness or affiliation a child might feel towards any given speaker, given the subjective nature of such an assessment.

Also, when asking parents to list individuals speaking with a different accent that their children had regular interaction with, I specify in the question whether I meant foreign or regional accents, as I wanted to capture exposure to both kinds of accents. However, most parents (26/37 responding yes) provided detailed information about where the individuals they listed were from, allowing me to break down the speakers listed into categories based on the type of accent (non-native vs. regional). Additionally, I had information from the second set of questions about bilingualism and speakers of English as a second language in the family, allowing me to compare their answers to these two sets of questions and determine if it was family members who speak English as a second language that were being referred to, or other community members.

When coding the data, I made the decision to not quantify exposure to a different accent, but instead made a binary categorization, either having exposure to another accent or not. This was in order to simplify the statistical analyses, as the range of responses would have either required creating overly broad categories, or smaller categories to capture all the responses, but with few subjects in each. In other words, the number of subjects would have been too small in any given response category to reach statistical significance (or the category so broad as to be meaningless). Thus in the following summary of exposure to regional accent and language history, the only answers recorded were "yes" (i.e. has exposure) or "no (i.e. has had no exposure to any variety of accent). Unfortunately, even using broader response categories, none of the variables representing exposure to accent influenced responses in any of the three tasks, with the exception of knowing another language.

Finally, I looked at overt awareness children may have of accents. In order to assess this without using an academic term such as "meta-linguistic" in the question, I asked parents

whether their child commented on people speaking with accents different than their own, and whether the child ever imitated these accents, and to cite specific examples.

2.4 Summary of Subject Demographics

The parents of 72 children signed up to participate in the study. Of the original participants, four chose not to participate, and one was excluded for having a speech IEP (Individualized Education Program), meaning that he had been assessed by the schools as needing speech therapy. One further child was excluded from the Experimental Task 2 analysis, as she informed me that she had a lisp, and "talked funny" like the Southern speaker in the experiment. This case is discussed in more detail in the description of the Experimental Task 2 results. This left 66 children who participated in the study.

In addition, five (different) children chose not to complete one of the two experimental tasks, so the total number of subjects completing each experimental task was only 61. Their data for the task they did complete was kept in the analysis. All of the children agreed to participate in the open-ended question task, even if they had not chosen to complete the experimental tasks, leaving 66 participants who completed this task.

Of the 66 subjects who took part in the study, 31 were male and 35 female. Thirty-eight were monolingual English speakers and neither Hispanic nor African-American. Thirteen subjects were bilinguals and 13 were African-American, but there were no bilingual African-Americans. Most of the bilinguals were of Hispanic heritage, but there were three from other backgrounds. Those three exceptions were: a child whose parents were from China; a child with a father from Italy who spoke Italian at home; and one child whose parents were Kannada speakers from India. Only three of the bilinguals were born outside of Northville, one each in New Jersey, Florida and South Carolina. Of the bilingual participants, seven were learning English as a second language, and six were simultaneous bilinguals, speaking English and one other language at home, as reported by the parents.

Nine of the total 66 participants had been born outside of the Philadelphia area, but only one of those had lived for more than half of his life outside of the Philadelphia area. That child had spent four of his six years in South Carolina, and is the same bilingual child born in South Carolina mentioned above.

The group of children with parents born outside of the Northville/Philadelphia area is larger, with 30 of the subjects having at least one parent born elsewhere. Sixteen of those had at least one parent born in a country other than the United States.

The average age of the children participating was 70 months, i.e. 5;10 years. The range of ages was 61 - 77 months. The distribution of ages is given in Figure 2.1. All subjects had been attending Kindergarten for four months when the study took place.

Most of the data on subject demographics was asked on the questionnaire: age, place of birth, gender, birth date, bilingual status. However, I did not specifically ask the parents or children to self-identify as far as race was concerned on the questionnaire, but rather recorded this information based on my own notes and any information the parents provided about their and the child's birth place. I included the child with a Chinese parent as Asian. Hispanics included all eight subjects with Mexican parents, as well as the children of families from Guatemala and Puerto Rico. I did not include the two children with both parents from Jamaica as African-American, but counted them in the "other" category with the subject whose family was from India. The child with one Jamaican parent however was kept in the African-American category because the parent identified himself on the questionnaire as African-American and commented on the child's African-American identity.

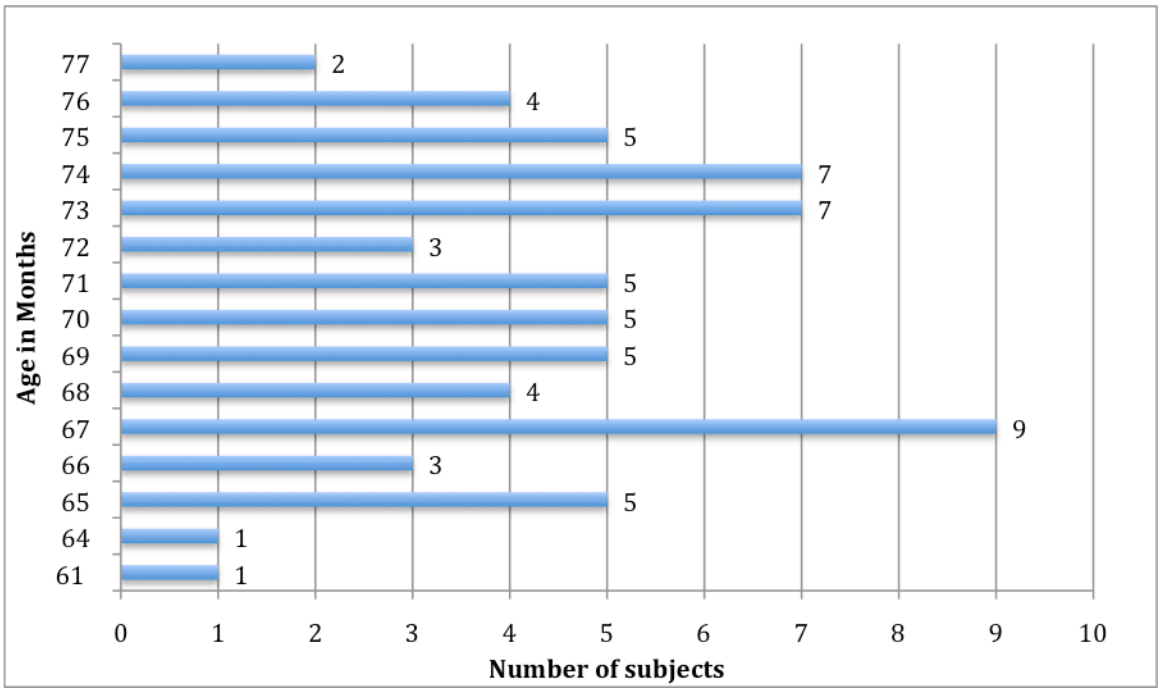

Figure 2.1 Distribution of Ages of Subjects (in months).

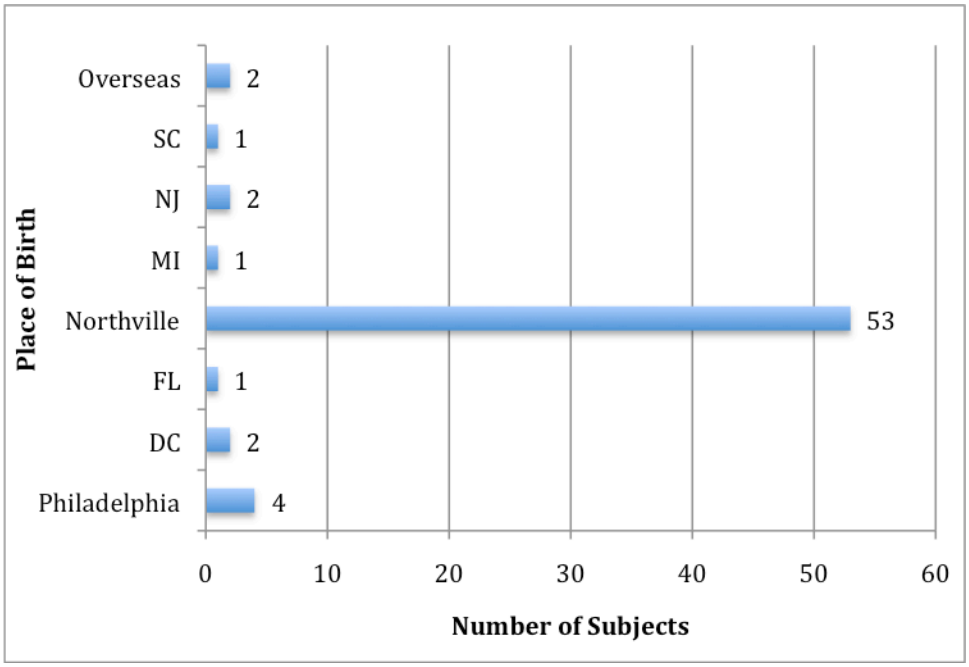

Figure 2.2 Subject's Place of Birth

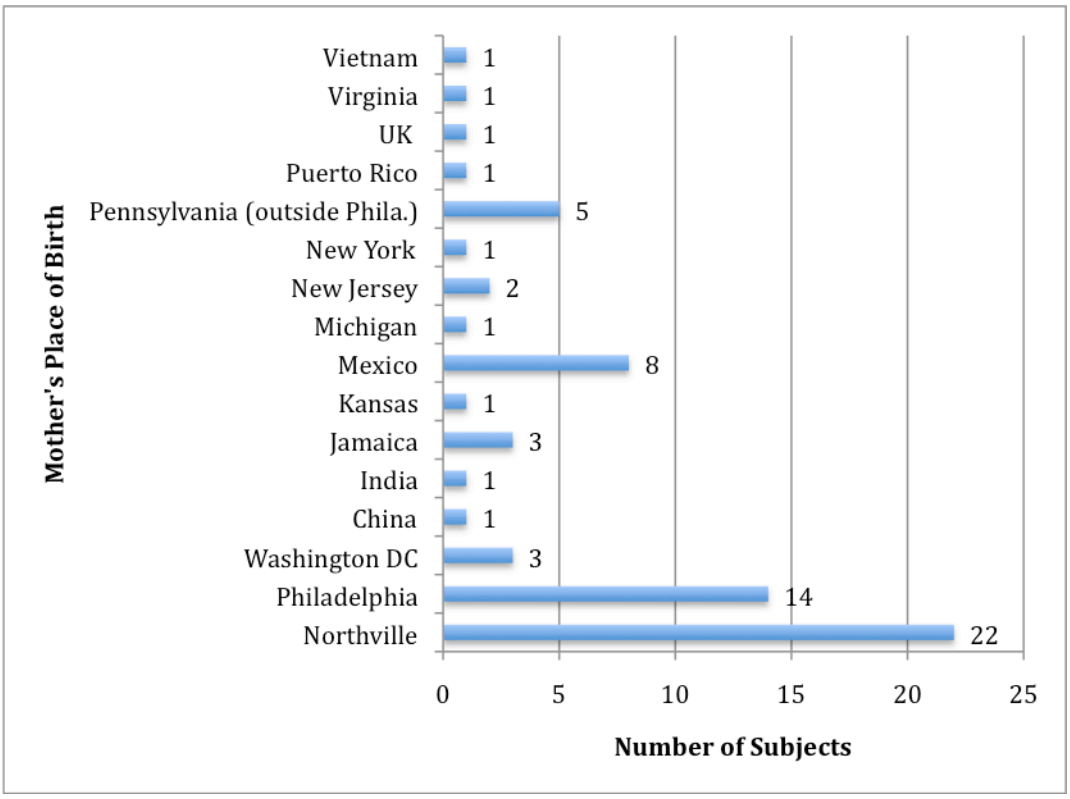

Figure 2.3 Mother's Place of Birth

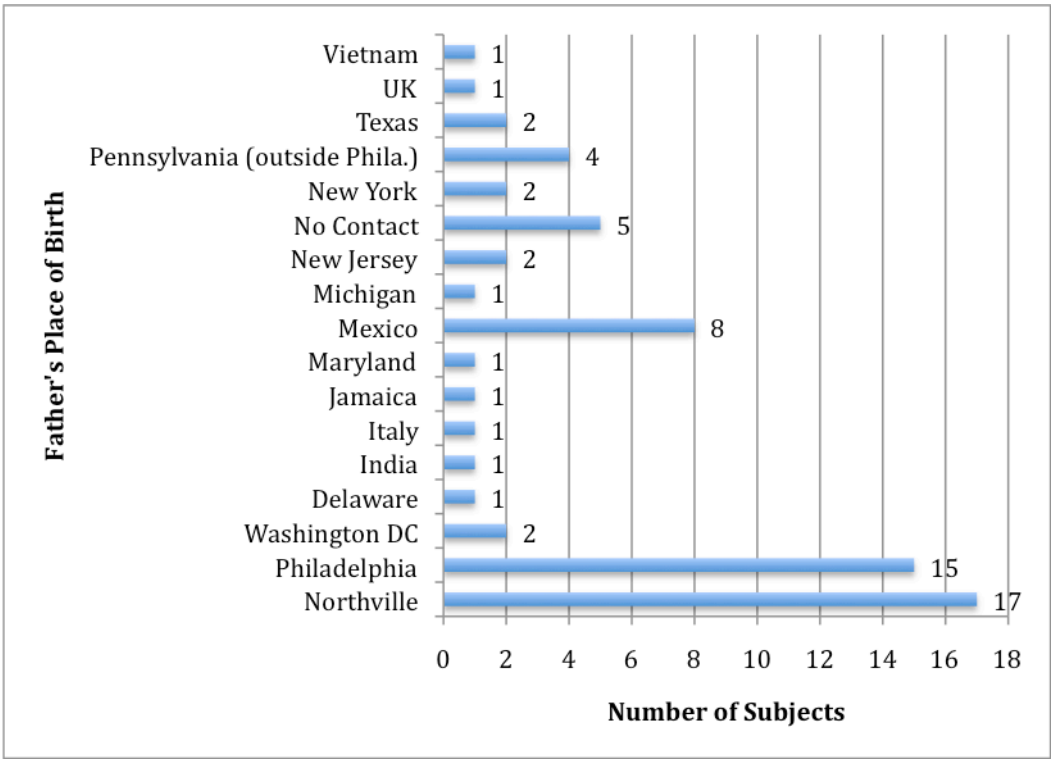

Figure 2.4 Father's Place of Birth

2.5 Summary of Language Background Data

The following table summarizes the results of the questionnaire not depicted in Figures 2.1 - 2.4 above.

Table 2.1 Summary of Accent Exposure Data for the Subject Group

	Travels regularly outside of area	Interacts with others with regional accents	Interacts with others with foreign accents	Exposure to accents on TV/media	Imitates other accents	Comments on other accents	Bilingual
Yes	33	8	29	22	34	34	13
No	33	29	29	44	32	22	53

As the numbers in the table show, there were fairly even distributions between the positive and negative responses for each question asked. I should caveat this chart by noting that

parents were likely not thinking about the school environment when responding to the questionnaire. All of the children participating in this study had regular exposure to fellow students who were non-native English speakers, some of whom had foreign accents in English, although few parents made note of this fact. In one school, where approximately half the subjects were enrolled, there was also a teachers' aide who moved between the kindergarten classes on a daily basis, and who spoke with a strong New England regional accent. None of the parents or students commented on her speech, although the children had daily exposure to her.

Given these oversights, it is likely these data aren't a completely accurate reflection of reality. However, they are an attempt at quantifying whether having any exposure to regional accents through different channels affects children's awareness of them, or the ability to report hearing or identifying accents.

The range of answers provided to the question about exposure to other accents in daily life was broad, and most accents were only mentioned once or twice across all of the questionnaires. However, several accents were mentioned multiple times: Spanish accent, 13; Southern accent, 3; and Jamaican, 3. The other accents mentioned were: New York, Iowan, Philadelphian, Texan, German, British, Korean, Indian, Chinese, Vietnamese, Japanese, New Jersey and one respondent mentioned an individual with a speech impediment.

Over half the subjects were reported to imitate or comment on others' accents, and the most oft-cited example was that the child would comment/imitate the accents of native Spanish speakers they heard. Few parents (8/37) reported children imitating regional accents, most frequently British and Australian accents heard on TV programs. Ten parents reported their children imitated the Spanish accents of their schoolmates. No other kinds of accents were mentioned in responses to the imitation question.

In response to the question about whether children ever commented on another accent, comments about Floridian accents were mentioned 13 times, British accents twice, and the child's own pronunciation as compared to someone else's pronunciation three times. Unfortunately few details were given, but these parents noted that the child commented on his or her own speech relative to another's at some point.

2.6 Conclusion

These data serve to illustrate the demographic composition of the subject population, as well as give an idea about the amount of exposure children in the subject population have to different regional accents. For this group, it appears that regular interaction with speakers having other accents is limited, and approximately half of the children do not comment upon regional accents or attempt to imitate them. This may be that they have such little exposure that they have no occasion to comment upon regional accents. It also could be though, that these accents are less salient to children of this age, and perhaps therefore do not receive commentary. The following section, which describes a task explicitly asking children about their knowledge of regional accents, provides further insight as to why children may not comment on accents. Furthermore, the different kinds of exposure to regional accents, as measured in the questionnaire, is tested as a predictive factor for whether children can correctly identify the source of regional accents when they are asked to do so.

CHAPTER 3

Awareness Task

3.1 Introduction

As explained in the previous chapters of this dissertation, I expect that kindergarten-aged children are capable of acoustically discriminating regional accents; that is to say they are capable of hearing the differences between phonetic features. However, children at this age may experience difficulty reporting that they have heard a regional accent. This difficulty may come as a result of not understanding the meta-linguistic aspects of regional accent. The task reported on in this chapter attempts to quantify and qualify the level of meta-linguistic awareness kindergarten-aged children have about regional accent.

The task was designed with two specific objectives. First, it enables me to get a descriptive picture of what these children know about regional accents, and how their understanding of it differs from adults'. Second, it allows me to quantify the level of meta-linguistic awareness of regional accent that the children have. This allows me to test whether this knowledge influences their performance on the experimental tasks, via statistical analyses.

This task is crucial for interpreting the results of the two experimental tasks in this dissertation, as many of the analyses performed in the following chapters examine the link between the results of the current task and of the experimental tasks. It also lays a foundation for further studies of how the meaning children assign regional accents changes over the course of language acquisition.

3.2 Background

Exemplar Theory is thought to explain the detailed social knowledge that hearing a particular accent or socio-indexical variable evokes for listeners. This phenomenon is attributed to the association of linguistic variables with social categories, created by repeated exposure to

persons identified with that social category. Thus, exposure, according to this theory, is the key factor in category formation. As tokens of speech associated with a social category are accumulated, they are thought to be abstracted, which allows other speakers exhibiting those linguistic variables to be identified as members of the social category. The more tokens that are accumulated, the better entrenched the category becomes, causing new input in the proximity of the category to be perceived as belonging to the category.

This mechanism works well to explain, for example, categorical perception; listeners do not perceive sounds that are acoustically midway between two established phonemes of their native language as an intermediate sound, but rather as belonging to one or the other phoneme category. Perhaps a better analogy would be gravitational pull: large, massive objects, such as planets, exert a stronger gravitational pull than a less massive object, such as an asteroid. A rock passing between the two would be drawn to the planet, not the asteroid, given the planet's superior mass. This roughly parallels how an ambiguous token would react in the presence of a "massive" or well-established category- it would likely be perceived as a member of that category and not as a member of a smaller, less established category in the same perceptual proximity.

The exact coordinate system of perceptual space is rather indeterminate in the literature. Johnson (1997) refers to F1 and F2 in the discussion of phoneme categorization in Exemplar Theory, and F0 in the discussion of identifying speaker gender (Johnson, 2006). However, entire words are not as easily described by F1 and F2, and non-linguistic social characteristics of speakers even less so. Thus, concepts of proximity and relative weight of tokens remain abstract, since no units with which to quantify them have been identified such that all kinds of variation can be accommodated.

Nonetheless, the crux of the theory remains that experience hearing tokens (of regional linguistic variation, in this case) builds perceptual categories, and the more experience with this kind of variation, the more robust the category. Thus, there is reason to believe that children with extensive exposure to regional linguistic variation might have a better ability to identify it than children with little exposure to regional variation. Furthermore, children who regularly encounter regional variation might also be more overtly aware of its existence. As discussed in the introductory chapter of this dissertation, but unaddressed in extant literature on the subject,

having socio-indexical information associated with categories of social variation may also play a facilitating role in identifying regional variation for young children.

Aside from the theoretical importance of establishing the role of exposure to regional variation, there is a practical, methodological reason to ascertain each subject's awareness and knowledge of regional variation. When designing a task to test awareness of regional accents, I have to first establish what children know about regional accents. It is not a reasonable assumption that children have the same understanding of regional accents as adults.

The questions for this task were created to address what I interpret as the three main assumptions about children's knowledge that have been made in prior regional accent studies. Those assumptions are:

1. Children are aware that there is organized regional variation in phonology.
2. Children recognize speech local to their community (and can identify non-local speech as well).
3. Relative geographic location isn't an abstract concept.

These three types of information comprise the basic underlying concepts one needs to interpret regional accents. Therefore, in creating the questions asked in this task, I tried to capture the children's understanding or state of knowledge about each of these three points, to see whether one in particular, or any combination of them, influenced their ability to report hearing regional accents in discrimination tasks. Below, I describe each in detail.

3.2.1 Children Know about Regional Variation

American adults recognize non-local accents with high levels of accuracy, and attribute the source of the accent to the region of the speaker's birth/residence, even if they are poor at identifying the specific location (Preston, 1986). The assumption is made in Preston's task that the subject knows about regional variation and can detect it. It may not be a safe assumption that the same is true of children.

It is possible that a child may attribute the source of difference to any number of personal qualities of a speaker (gender, age, etc.) or simply as an idiosyncratic speech style. Another possibility is that the child may not even try to imagine a reason that people speak differently, and never have given the source of difference any consideration, and won't, until this difference acquires a social significance to the child. Therefore, using the term "accent" when giving

children instructions in these experiments would be at best, meaningless, and at worst, confusing, to a child. As a result, I could not base any task on explicit references to accent, but had to make the reference point something that all subjects could identify and relate to.

3.2.2 Children Understand There is a Local Variety of Speech

Many children grow up hearing a wide range of kinds of speech: regional varieties, ethnolects, varieties linked to social status, to name a few. Given the amount of variation that children are exposed to in speech, assuming that a kindergarten-aged child has determined there is a "typical" variety for their area may be a fatal assumption for an accent discrimination task. They may have yet to understand what is typical and atypical for their community, and what the sources of all different kinds of variation are. By asking whether they are able to identify a local regional variety and distinguish it from a non-local variety, I can at least determine whether they have made the generalization that certain kinds of speech are commonly heard in their environment.

3.2.3 Children Understand Relative Geographic Location

Perhaps the most important element for interpreting regional variation is understanding that there are different geographic locations, separated spatially, politically and culturally. Five year olds generally are not adept with the geographic and spatial concepts that regional accents are built upon. Children have been shown to first acquire the concept of nationality around the age of five (Anderson, 1990) along with an expectation that people from different places might act or speak differently than they (Hirschfeld & Gelman, 1997). "Nation" often encompasses (or is described as encompassing) differences in language and culture that are comprehensible or plainly audible (or visible). The more subtle distinction between regions within a nation may not be obvious to a five year old, although this has not been tested to my knowledge. Thus, even if a child has regularly traveled to a place where everyone speaks with a regional accent other than the child's, that child may not identify the place as physically separate or distinct from the home region.

3.2.4 Factors Influencing Meta-linguistic Awareness of Accent

The three areas of knowledge described above likely emerge as children are exposed to different varieties of speech, and overt talk about accents, and stereotypes of language, or "folk linguistic" concepts, as described by Preston and Niedzielski (1999). I expect that the subjects will have varying levels of awareness about regional accent, depending on their age, exposure to other regional accents, proclivities for language, etc. For this reason, factors contributing to the development of awareness must also be taken into account in order to explain individual differences in performance as well.

Given my hypothesis that part of being able to recognize and report hearing a regional accent depends on awareness of the existence of regional accents, which in turn may only develop with repeated exposure to regional variation, I look at whether any of the measures of exposure to regional accent described in the previous chapter, as well as demographic factors such as age, race and gender, predict responses to the questions in this task.

I don't expect the amount of exposure to, or awareness of, regional accents to be uniform over the subject population. Therefore, instead of considering group performance on discrimination tasks, I use the results of the awareness questions in this task, as well as individual measures of exposure to regional accents, to find correlations between them for each subject in the experimental tasks. Analyzing the data for individuals provides a more nuanced picture of the relation between the three elements of this study: regional accent discrimination, awareness and exposure.

3.3 Methods

This task is a series of five open-ended questions was the final task in a three-task study of kindergarten-aged children's ability to discriminate, identify and interpret regional accents.

3.3.1 Stimuli

Stimulus materials included five questions posed orally by the experimenter, a map of the United States and four audio clips of speakers representing the two different regional accents (Philadelphia and General Southern) used in all of the tasks reported on here. The audio clips were taken from recordings of word lists read by six native adult speakers of English (three from each of the two regions) who were life-long residents of their respective home areas. The words

used in the present task to exemplify the two accents were from the same list of words heard in the two experimental tasks (For a more detailed description of the recorded word lists, please see the Stimuli section of the following chapter, as well as Appendix 2). The stimulus words for these tasks were chosen and recorded expressly for this study to highlight differences between Southern and Philadelphian vowels in speech.

The five questions used in this task were as follows:

Question 1: Can you point to where we live (*while looking at map of U.S.*)?

Question 2: Can you show me and name any other places you know (*while looking at map of U.S.*)?

Question 3: (*After hearing a short clip the local speaker*) Does this person sound like he lives here? (*if answer is no, have the child say where person in from)*

Question 4: (*After hearing a short clip the non- local speaker*) Does this person sound like he lives here? (*if answer is no, have the child say where person in from)*

Question 5: Can you guess why these two people talk differently?

These five questions were designed to capture the three main pieces of knowledge needed to interpret regional accents: a) knowledge of one's position vis-a-vis other groups who might speak differently; b) the expectation that there is a typical or commonly heard variety of speech in their environment, and people speaking with a completely different accent are not from the area, and finally, c) that regional accent is tied to geography and not some other source of variation.

My purpose in designing discrete questions whose answers could be evaluated as correct or incorrect was to attempt to quantify the children's awareness. The questions were either scored as correct (1) or incorrect (0). By quantifying them, this allowed me to use the scores from individual questions, as well as the total number correct out of the five questions, as an independent variable in statistical analyses.

In addition, I also recorded any descriptive answers children provided to these questions, in order to capture how they interpret regional accents, especially where their interpretation deviates from an adult-like understanding. These comments provide insight into the wide range of interpretations individual children have of regional accents.

3.3.2 Participants

All 66 children participated in this task, regardless of whether they completed only one or both of the experimental tasks. For a detailed description of the subject population, see the previous chapter.

3.3.3 Procedure

The five questions were posed in the same order for each subject in the Awareness Task. The questions were ordered so that subjects were not biased in their responses by previous questions. The Awareness Task also always followed the two experimental tasks. This, of course, means that the experimental tasks could have influenced the likelihood that children provided answers based on the knowledge that the two previous tasks were about regional accents. However, I thought this to be less problematic than to explicitly ask children questions referencing geography and accent before participating in the experimental tasks. Because I was particularly interested in whether children identified regional accent when not given any directions to do so in the experimental task, and therefore kept all of the experiments in a fixed order.

Another justification for maintaining the order of the experiments is that if children had provided a reason other than regional accent that the two groups of speakers spoke differently in the Awareness Task, this might have primed them to listen for that particular difference in the two experimental tasks. This would have skewed the results of the experimental tasks. If the reverse were true, and the children divined some other speech quality by which to discriminate speakers during the experimental tasks, and were thus primed to give that as an answer when asked why the speakers talked differently, this would affect only one of the questions in the Awareness Task, and not skew the results of the two experimental tasks.

3.3.4 Predictions

The research questions and predictions relating to this task are repeated for convenience below. These are taken from the general set of questions and predictions for this study outlined in Chapter 1.

Q2: Are children able to use the socio-indexical information encoded in a regional accent to discriminate between speakers?

H2: Five to seven year olds will generally be able to identify members of their local community based on accent.

Q3: Is ability to discriminate between regional accents dependent on having direct experience with the regional accents in question?

H3: Given that Exemplar Theory has emphasized the role of storing and abstracting tokens of speech in the creation of categories, I would expect that children who have direct experience with the non-local accent to be superior in identification of regional accents as compared to children without that experience. However, children with experience with other non-local regional varieties may be able to extrapolate from this knowledge to identify a novel accent as non-local.

All children in this study will have had intensive exposure to the local regional accent represented in the stimuli. However, I may find differences are in children whose parents are not originally from the town where this research was conducted, or children whose parents are speakers of another variety of English (i.e. African-American English, L2 English, etc.) or another language entirely. These children, although they will have heard the local variety extensively in school, will have had most of their exposure prior to starting school in another variety. Therefore, their categories for local speech may be defined differently than those children whose families speak the local variety, and this in turn may influence the correlations between experience and discrimination of the regional varieties. It would likely not improve their discrimination, since the variation they are most familiar with is not regional variation.

Q4: Is ability to discriminate affected by children's overt awareness (i.e. ability to identify a regional accent and state what the regional accent indexes) of the accents in question?

H4: Given that there are effects of stereotypes of speakers with an accent in perception studies on adults, it would be surprising if children who have an overt knowledge about accents did not also use this information when making decisions about who sounds similar. However, I expect relatively few children of this age to have an overt general knowledge about regional accents and their distribution, as they may lack the geographical awareness to comprehend relative distances and location, making geographically-based variation uninterpretable to them.

The above questions and hypotheses relate to the entire study comprised of three experiments; with respect to the specific questions asked in this task, I have the following hypotheses:

Question 1: Can you show where we live (*while looking at map of U.S.*)?

Question 2: Can you show me and name any other places you know (*while looking at map of U.S.*)?

Hypothesis for Questions 1 & 2: I expect most children to know where they are located, but will be less able to identify other places on a map, as I don't expect most children to be conscious of the relative location of other places, even if they have heard of or visited other states/cities.

Question 3: (*After hearing a short clip of each speaker*) Does this person sound like he lives here?

Question 4: Can you guess where this person is from (*if either speaker is identified as not living in the area*)?

Hypothesis for Questions 3 &4: I expect that a majority of children will be able to identify the Philadelphia speakers as sounding local and the Southern speakers as non-local. However, I don't expect that most children will be able to identify the Southern speaker as being from the South, although I do expect that at least some children will be able to say Southern speakers are from far away, showing a nascent linkage between regional accent and geographical distance. If this prediction is borne out, it would show that children can use accent to distinguish between community members and outsiders, as defined by geography.

Question 5: Can you guess why these two people talk differently?

Hypothesis for Question 5: I don't expect that most children will be able to say that the speakers talk differently because they are from different places, or that accent varies by region. Although this question seems to ask the same information as Questions 3 and 4, i.e. asking them to locate the speakers and make a statement whether each accent is local, Q5 requires them to abstract the differences and make a general statement about the relationship between geography and regional accent, which would require a high level of meta-linguistic awareness.

3.4 Results

The five questions asked of the children were scored as 1 (correct) or 0 (incorrect). For the purposes of conducting statistical analyses, if no answer was provided, it was scored as incorrect. However, I also describe some of the alternate answers subjects provided to see the information these answers provide about children's state of knowledge of accents.

The two tables below show the raw scores for each question and average scores for different sub-groups. As can be seen in the table, the subjects were best at identifying their home on a map of the United States (Q1), and also did well in indicating which speaker did not sound

like he was from their hometown (Q3). However, they were not able most of the time to indicate where the Southern speaker was from (Q4) or identify any other places on a map of the United States (Q2). Q5 was the least often answered question, since as expected, most subjects had difficulty abstracting about the distribution of regional accents.

The results of the responses are summarized here:

Table 3.1 Summary of Responses to Awareness Task (n=66)

	% subjects responding correctly	Raw #	# (%) not providing answer
Q1	70%	46	2 (0.03%)
Q2	50%	33	33 (50%)
Q3	65%	43	2 (0.03%)
Q4	42%	28	2 (0.03%)
Q5	32%	21	28 (42%)
Average Total Awareness Score: 2.40/5			

Included in the next table are the results of several sub-groups of subjects, including Monolinguals, children with both parents born outside the area where the study was conducted ("Outsiders"), children who correctly answered Q5 (i.e. children who could state that accent of the speakers was linked to region) and children who provided a reason other than geography for the difference in accent between speakers for Q5 ("Other Theory"). The total average Awareness Score for each of these sub-groups is reported in the last row of the table. The significance of the results of these sub-groups will be discussed in subsequent sections.

Table 3.2 Comparison of Sub-group Results on Awareness Task

	Q1	Q2	Q3	Q4	Q5	Total Score
Q5 Correct Answer (n=20)	75% (15)	60% (12)	85% (17)	80% (16)	100%	2.38
Q5 Other Theory (n=16)	63% (10)	50% (8)	38% (6)	19% (3)	0	1.42
Monolinguals (n=53)	68% (36)	55% (29)	72% (38)	51% (27)	40% (21)	2.75
Bilinguals (n=13)	77% (10)	31% (4)	38% (5)	8% (1)	0% (0)	1.46
Outsiders (monolingual) (n=13)	85% (11)	69% (9)	54% (7)	31% (4)	38% (5)	2.62
Insiders (n=41)	61% (25)	49% (20)	76% (31)	56% (23)	39% (16)	2.73
African-Americans (n=15)	53% (8)	40% (6)	60% (9)	47% (7)	33% (5)	2.23
Caucasian (n=38)	71% (27)	61% (23)	74% (28)	53% (20)	36% (19)	2.89
Other Race or Ethnicity (n=12)	83% (10)	25% (3)	42% (5)	8% (1)	8% (1)	1.54

The first two questions had the smallest range of answers provided: many children could point to where they were from, and the rest simply didn't provide an answer. I required children to both point and name locations to be counted as correct, since simply naming the town was a piece of information that could be memorized and didn't indicate anything about their geographical knowledge or sense of relative location. More than two-thirds (70%) of the

subjects could point out where they lived, but only half could locate another place besides their home on a map.

The third and fourth questions asked the subjects to say whether children could identify the local regional accent, and the Southern accent as non-local. Sixty-five percent of children could identify the local speaker as being from their community, and 42% correctly identified the Southern speakers as not being from the community. Thirteen children responded that both speakers sounded like they were local, and two subjects reversed the local and non-local speakers.

The children were asked to indicate where the speakers they identified in Q3 and Q4 as local and non-local were from, if their initial answers were incorrect. This served the purpose of determining if the children associated those accents with other localities, or were guessing at answers. Nineteen subjects provided an alternative to the correct answer to Q3. For the 12 subjects who said both speakers were local in Q3 and Q4, all indicated that both speakers were local, reconfirming their initial responses. For the three who said local speakers were non-local speakers in Q3, they all said the Southern speaker was from the Northville area. None of the three could guess where the actual local speaker was from. One subject said neither speaker was from Northville but also didn't provide a place where they could be from. Three subjects didn't answer the question.

In Q4, I did not expect that most of the children would correctly locate the Southern speakers as from the South, as only seven of the participants had any prior interaction with southern-accented speakers, based on the results of the parent questionnaire. I marked any answer indicating that the speaker was from some clearly different location, whether they pointed on the map, or if they simply said "from another state/country" or provided the name of another state or country. I treated responses to the question differently than in Q2, not requiring children to point to *and* name a specific place on a map in order to count the response as correct. This is because there was no chance children could have memorized an answer to this question, and I was simply interested in whether they'd designate the speaker as not being from their home region, even if they had no idea where the region they named is located, or whether they'd place the Southern speaker as coming from somewhere far away from their home on the map. I later checked to make sure that children who pointed to some distant location on the map in response to Q4 had gotten Q1 (Where do you live?) correct. Had they not known where they lived,

pointing to another region on a map would have been hard to interpret. All of the children who pointed to another area however also knew where they lived and could point it out on a U.S. map.

I did not come across cases where the children designated the non-local speaker as being from a neighboring town or region, and did not have to interpret whether they perceive a neighboring region as having a different speech variety, or whether they perceive that community as being very distant, and therefore a possible home of an atypical sounding speaker, regardless of the linguistic reality of speech in that place.

There was a broad range of places that children said the Southern speakers hailed from. California, the Dakotas and Nebraska got picked as probable homelands for the Southerner in three, four and two subjects' responses respectively, probably because of their distance from Pennsylvania on a map. Florida and Kentucky were each chosen once. None of the kids choosing these areas named those states, save one who told me "They talk funny in California" after providing his response. However, Texas was explicitly named as the home of the Southerners by five different subjects, and a 6th subject guessed the Southern speakers were "from a desert, because they sound like a cowboy." Two additional subjects guessed the speakers were from another country.

Several subjects had very detailed responses to Q4. One told me the Southern speaker was from Kansas, as she had visited there and heard similar sounding speech. Another child told me he had heard someone talking "like that" (referring to the Southern speakers) at the beach in New Jersey, but didn't think the person was from New Jersey, and declined to guess where the speaker may have been from. These two responses are intriguing, as it is clear these children paid particularly close attention to how people speak in different locations and were able to recall where they had heard similar speech.

The fifth question, Q5, was the most telling about the general connection of regional accents to geography. Twenty subjects provided the correct answer. An additional 18 provided an incorrect answer, attributing the differences between the two regional accents to differences in pitch of voice, individual variation or gender (despite the fact all of the speakers were of the same gender). The rest of the subjects, n=28, declined to provide an answer. Perhaps unsurprisingly, none of the 15 subjects who incorrectly identified the Southern speakers as being

local in Q3 gave a correct answer to this question (nine provided a reason other than geography, and six didn't respond).

3.5 Analyses

In addition to tallying the responses of different groups, I also conducted a statistical analysis to see if any demographic factors, as measured by responses to the Parent Questionnaire, had an effect on the likelihood that any given subject answered a particular question correctly. I first totaled the number of questions in this task that the subject answered correctly and plotted this score against several candidate variables from the Questionnaire. This allowed for a visual inspection of the data to find possible correlations between awareness and other variables. The most plausible variables were age and having one or more parents from outside the Philadelphia area. Interestingly, none of the variables relating to exposure to regional accents (such as travel, interaction with individuals with other regional accents, exposure to media where regional accents are portrayed) seemed to have any effect on subjects' awareness of accent, either individually or as an aggregate measure. Race, which I suspected might have an influence on perception of accent, also did not play a significant role in predicting results.

I used a general linear model to test whether age and having parents from outside the area correlated with better performance on each of the Awareness questions, and then the aggregate score for all five Awareness questions. The effects are summarized in the tables below:

Table 3.3 Question 1: Can You Find and Name Where You Live on a Map?

	Estimate	Standard Error	T Value	Pr(>|t|)
Age*	0.02917	0.01440	2.025	0.0471
Outsider*	0.15328	0.06564	2.335	0.0227

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

For Awareness Task Question 1, increased age and having parents from outside the region in which the child lives positively correlated with answering this question correctly.

Table 3.4 Question 2: Can You Find and Name any Other Place on a Map?

	Estimate	Standard Error	T Value	Pr(>|t|)
Age	0.017686	0.016606	1.065	0.291
Outsider	0.004921	0.075677	0.065	0.948

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

For Awareness Task Question 2, neither factor correlated with answering this question correctly.

Table 3.5 Question 3: Does this Person Sound Like He Lives Here? (Local Speaker)

	Estimate	Standard Error	T Value	Pr(>|t|)
Age	0.009886	0.014865	0.665	0.50843
Outsider**	-0.202895	0.067740	-2.995	0.00392

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The negative T value for the correlation with being an Outsider indicates that children with both parents from outside the Northville area were less likely to answer Q3 correctly than those with at least one parent from the Northville area.

Table 3.6 Question 4: Does this Person Sound Like He Lives Here (Non-local Speaker)

	Estimate	Standard Error	T Value	Pr(>|t|)
Age	0.01771	0.01559	1.136	0.2602
Outsider*	-0.18056	0.07103	-2.542	0.0135

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

For Awareness Task Question 4, having parents from outside the region in which the child lives negatively correlated with answering this question correctly. This means that children whose parents were not natives of the child's hometown were less likely to identify the non-local speakers heard in the stimuli as non-local.

Table 3.7 Question 5: Why do These Two Speakers Talk Differently?

	Estimate	Standard Error	T Value	Pr(>|t|)
Age***	0.05157	0.01347	3.829	0.000299
Outsider*	-0.14694	0.06138	-2.394	0.019664

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Again, in this question, there is a negative correlation between being an Outsider and being able to identify regional variation as the reason the two sets of stimuli speakers sound different. However, being older has a positive correlation with answering this question correctly.

Table 3.8 Factors Correlating with Awareness Score

	Estimate	Standard Error	T Value	Pr(>|t|)
Age**	0.13031	0.04821	2.703	0.00883
Outsider.	-0.37100	0.21972	-1.689	0.09626

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Looking at the aggregate awareness score, older children were more likely to answer questions correctly than younger children.

For Question 1, Question 5 and the aggregate Awareness Score, age has a significant effect on whether the subject answered correctly (note: in the case of Question 5, I refer to any answer dealing with geography as correct. This is not to say that some of the other answers were not correct, such as those saying differences were due to individual variation, or pitch of the speakers' voices. However, for the sake of brevity, I will refer to "correctness.") Older children were more likely to answer these questions correctly, and have higher overall awareness scores.

There were some very interesting results vis-a-vis the birthplace of the parents and the children's responses to this task. For the questions asking children whether a given speaker sounded local, and why the Philadelphian and Southern speakers sounded different, having two

parents not originally from the Philadelphia area decreased the likelihood that those children would provide a correct answer (indicated by the negative t values in the tables). In other words, having intense exposure to another regional variety complicated the associations children make between a local regional accent and a single geographical location. Having at least one locally born parent increased the likelihood that children would correctly identify local and non-local speakers, as well as give geography as a reason that people spoke differently.

3.6 Discussion

I designed this task to test children's knowledge of three main concepts underlying the ability to correctly interpret regional accent: understanding of geographic locations, understanding that local and non-local speech varieties exist, and concept of speech varying generally by region. My hypothesis that children have an emergent concept of regional accent is supported: the majority of children know there is a local speech variety, and 42% can identify an unfamiliar regional accent as non-local. However, they have less success generalizing this knowledge to make a statement about geographical distributions of regional accents. Below I discuss the findings for each of the three underlying conceptual areas I addressed in this task, and what role they play in the development of the concept of regional accent.

3.6.1 Geography and Spatial Concepts

I attempted to test whether knowledge of relative geographic location influenced the ability to report hearing a regional accent, or correctly interpret a regional accent. Approximately half the subjects could identify another place on the map relative to their home (Q2), and two thirds of the subjects could find their own homes on the map (Q1). Being older and having parents from outside the region seems to correlate with better performance on Q1. This was the expected correlation; older children likely are better at understanding maps and geographical concepts, and children who are aware of geography thanks to familial connections had an advantage in identifying where one lives.

These two factors didn't correlate with success in Q2, however. This is puzzling, as I expected the same factors to correlate with correct answers to this question as well, and 28/66 subjects answered this question correctly.

One potential problem with these questions may be that they are only testing rote learning and not any deeper understanding of geography and relative location. A better, more in-depth test of understanding relative distance and geography should be designed for future studies.

3.6.2 Local vs. Non-local speech

As predicted, a majority of children could identify speech as local or non-local (Q3). They also were highly skilled at saying the speaker they identified as local was from their hometown. However, identifying where non-local speakers was from proved more difficult. This is perhaps unsurprising, as this requires the experience of hearing a set of particular set of phonetic features and associating them with a particular location in the country. As past research on adults has established (Clopper, 2010; Preston, 1986) most adults have only inexact ideas of which phonetic features constitute an accent from a particular region, and the regions adults associate with an accent is subjectively determined by each individual. So it is unsurprising that children with less exposure to different regional accents could correctly identify exactly where an unknown speaker was from. However, it is clear that some children understood that difference in speech marked a speaker's place of residence.

The difference in the number of children who could recognize and locate local speech vs. the number who could do the same for non-local speech is nicely illustrative of how the concept and categories of regional accent develops with experience hearing it. Most children in this study have established a category for local accent and correctly associate it with their home region. However, there is more variability in how the non-local accent is interpreted. Some recognize the unfamiliar accent as non-local. Some have a more refined "non-local" category, with points on their mental dialect map slowly resolving: six children named Texas as a probable location of the Southern speaker, so they must have begun to develop a category for Southern, or at least Texan, speech. However, a majority, 60%, was unwilling to identify the unfamiliar accent as non-local, much less locate the speakers in a specific geographic region. This may indicate that they were, for some reason, unwilling to assert that the source of variation was geographic origin, or that they did not know this was the kind of variation they were hearing.

A possible reason some kindergarten-aged children are unwilling to say speakers with unfamiliar accents are non-local is that they have heard similar-sounding speech (or some other regional variety distinct from the local varieties) in their hometown, and are thus unwilling to

generalize about the kinds of speech they hear in their environment. In other words, some children may be aware of a great range of speech variation in their environment, making them unwilling to rule out the possibility that a novel variety can't be found in their home region.

Evidence for a hyper-awareness of variation in the environment is found in the results of the Outsider subjects in this study. Thirteen of 66 subjects had two parents from outside the area and were monolingual. Counting both Bilinguals and Monolinguals, 32 of the 66 subjects had at least one parent from outside the area, a significant portion (49%) of the subject population. Of the 34 children who answered both Q4 and Q5 incorrectly, 21 had at least one parent from outside the Philadelphia area, and of those, five had both parents originating outside the area. Since these subjects regularly hear two or more regional accents from speakers who are "local" to them, the connection between place and regional accent may take more effort or exposure to establish.

It is worth noting however that few of the parents were from the South, so for most subjects, the Southern accent was unlike what they hear at home. Although few Outsiders had experience with the non-local accent used in this study, their linguistic experience might make the connection between accent and location unclear, or at least provide contradicting evidence to the idea that people in their area speak similarly.

I expect that once children are older and understand that place of residence isn't an immutable trait, having parents who speak with different regional accents facilitates reporting differences in accents. Indeed, the findings of Floccia, Girard et al (2009) support this: seven year olds with a parent from outside the subjects' home region did better at regional accent discrimination than those with locally-born parents.

One result is particularly notable by its absence from the range of answers provided to Q3 and Q4: none of the children likened the non-local speech to categories such as race or speaking another language. Given the racial and linguistic diversity of the community in which I conducted the study, I expected that at least some children would associate different sounding speech to ethnicity or race. Not one subject mentioned these as possible factors for the difference.

3.6.3 Generalizing the Connection Between Location and Regional Accent

Only 34% of children provided an answer related to geographic location for why the two sets of speakers sounded different. Some children who did not answer Q5 correctly (or failed to provide an answer) correctly identified the local and non-local speakers in Q3 and Q4. These apparently contradictory cases may simply be a result of inability to formulate a generalization, or alternately, not being overtly aware of the general geography-language connection, although they could comment on a case specific to their own environment. The responses I collected do not allow me to distinguish between these two cases, although in either case, it is clear that experimenters should not assume that asking children to sort accents by location is a logical exercise in the eyes of their subjects.

For Q5, age positively influenced ability to generalize the regional accent-geography connection. Older children possibly are more aware of regional variation, and better at generalizing statements about the connection of language and location. This fits with the findings of Girard, Floccia and Butler (2008; 2009). In these studies, as described above, it was not until seven years of age that children could group speakers by regional accent when told that the speakers were from two different locations. The five year old groups failed at the task.

In order to illustrate how kindergarten-aged children do interpret accent, I examined the range of responses provided to Q5 that were not related to geography. Below I paraphrase the alternate explanations provided:

"One speaker doesn't have a normal voice"

"One speaker has a higher voice"

"They are men" (*3 subjects provided this answer*)

"People are born different"

"They don't sound like us"

"They are different people" (*5 subjects*)

"People have different voices" (*3 subjects*)

"They sound the same" (*3 subjects*)

These responses can be grouped into several different categories, which I summarize in the chart below:

Table 3.9 Summary of Incorrect Responses to Q5

Reason	Number of Responses
Gender	3
Individual Variation	9
Pitch/Voice Quality	2
Unspecified Difference	1

It is not clear why three subjects identified gender as the reason for the difference, as all the speakers of the stimuli were male. However, for these subjects, it is clear that the most salient difference between the speakers was not regional accent. Again, this does not mean the children did not hear the regional differences, but they were not the most obvious or important difference between the speakers, or the children were unable to formulate a general statement about regional accents. Supporting this idea is the fact that two of the children who gave individual variation as an explanation also provided, after considering for a moment, geography as a secondary response. I believe this indicates that they are aware that there are multiple possible sources of variation in accents, even if they cannot correctly or immediately identify them.

It is also interesting that the three subjects who reported that both groups of speakers sound the same performed above chance on the experimental tasks, meaning that although they were not able to operationalize the concept of accent or identify accent-based differences, they were actually able to hear the acoustic differences.

It is in cases like these where the distinction between *perceiving* and *reporting* hearing regional accents must be made. As the following chapters will show, the children were able to discriminate between the two accents with a high rate of accuracy. However, in this task, where they must incorporate socio-indexical information about regional accents, the answers they provide don't indicate a failure to discriminate, but rather missing socio-indexical information about the accent. This is likely the result of non-existent or emergent exemplar categories for the Southern accent heard in this experiment. Because it is a novel accent for most subjects, they have no category for it and as a result, there is no socio-indexical information about the speakers associated with it. It seems that at this age children can't reason that a novel regional accent

must be non-local, perhaps because the general concept of regional accent has yet to develop for most five to seven year old children.

It is also worth pointing out that despite the fact all of the subjects had regular exposure to multiple varieties of speech in their school environment (African-American English (AAE), Spanish, Spanish-accented English) and that over half had exposure to accents either through travel or interaction with speakers with regional accents at home, the majority were unable to state the connection between accent and location.

There are several possible reasons that exposure to other accents of any kind does not seem to influence identification of regional accents. First, before the age of six, any travel hasn't been of long enough duration to have an effect, or children weren't cognizant of regional accents they may have encountered during their travels.

Second, in order to for children to know that someone in their environment has a regional accent, they perhaps also have to be aware that the person is not originally from the same area. As seen in the responses of children who have Outsider parents, exposure to non-local varieties at home may complicate the ability to differentiate local and non-local accents. In these cases, the speaker is clearly local to the child, and so the connection between the locale and the accent is not apparent. This explanation is supported by research conducted by Chambers (2002), who reported that children of parents with non-native accents were often unaware that the parent had a foreign accent until late childhood. Although Chambers posits a kind of "accent filter" to explain this phenomenon, I believe it is due to an unclear connection between locale and language for Outsiders. It is clearly apparent to the child that the parent with the non-local accent is local from the child's perspective, and as a result the child does not know the accent is regionally motivated. This would be unsurprising, given that listeners can be induced to hear accents in speakers that have none when they believe the person is a foreigner or belongs to a particular speech community (Hay & Drager, 2010; Hay, Warren, et al., 2006; McGowan, 2011; Rubin, 1992). This phenomenon, if substantiated, would be that effect in reverse.

3.6.4 Theoretical Issues

I had thought that children with exposure to other regional accents might be able to reason about novel regional accents, to correctly interpret a non-local regional accent as such. However, the results of this task don't show a positive correlation between experience hearing

accents and accent identification, or experience and overt knowledge of regional accents. Instead, children who might be predicted to be the most skilled at recognizing and identifying different regional accents based on their exposure in the home to regional accents, are in some ways less aware, or less able to articulate their awareness, than children who have exposure to only one regional accent on a regular basis.

Outsiders, the children in this group with the greatest time and intensity of exposure, show negative effects of this exposure on their ability to correctly answer most of the Awareness Task questions. This is likely the result of the circumstances of their exposure, and not the frequency or intensity of the exposure. The negative correlation for the Outsiders may provide some insight on two of the open questions about the details of Exemplar Theory.

First, the idea that children's knowledge of social variation and social categories is strongly linked to individuals appears to be supported. The problem in the case of the Outsiders is that many Outsiders are not yet aware of the regional distinction between their parents and other members of the local community. Their inability to draw on that socio-indexical information may be attributed to the fact that they have the same indexical information linked to multiple accents, or that because the information they have accumulated is conflicting, it perhaps hasn't abstracted beyond the stage where their representations consist of individual speakers. In either case, the conflicting information inhibits use of socio-indices to assist in discrimination for the Outsiders.

By all accounts, Outsiders should have at least two robust categories for the two (or more) kinds of regional variation they hear on a daily basis, yet they still are unable to identify the sources of variation or correctly identify the speakers as local or non-local. This may be the result of hearing multiple accents from speakers who, for all intents and purposes, are local to the children. Once they correctly label these speakers as local and non-local, I suspect they will be able to generalize and identify speakers with other regional accents as being non-local. This remains to be tested, but these results seem to underscore the importance of overtly recognized category labels in correctly identifying sources of variation.

3.7 Conclusion

These results of the Awareness Task show that it is prudent to ask whether children understand a novel regional accent is from another a region, or even represents regional variation

when designing regional accent discrimination tasks. Most of the subjects of this study did not have the ability to identify a novel regional variety, or even recognize it as regional variation. This is not to say that kindergarten-aged children are deaf to the differences between regional accents. The difficulty lies in knowing what they are hearing is regional variation and what it signifies about the speaker. Based on the results of this task, it also appears that designing the two experimental tasks with no reference to accent or abstract reference to location was justified, and may be the source of difficulties experiences in other tasks conducted with early-school-aged children.

Subjects demonstrated the ability to recognize and identify local speakers with a high rate of accuracy, even when their primary source of input is not in the majority regional variety of their home region. However, they are less successful reporting that speakers with an unfamiliar accent are non-local, or even identifying the non-local accent as regional variation. This is likely the result of the varying levels of experience with the local accent and non-local accents, and a lack of information about what the accent signifies about the speaker. Ability to identify regional accents, or even the existence of regional variation, is dependent on direct experience with speakers using those regional varieties. It is therefore unsurprising that they are very aware of the local accent, and less able to say anything concrete about speakers with a novel accent.

However, the kind of exposure children get to regional accents determines how they interpret it. Although Outsiders have exposure to two regional varieties in their environment, the circumstances of that exposure complicate their understanding of how regional accent is distributed. I suspect that because they hear two regional accents locally, they do not understand that the accents have to do with region, and may attribute the variation to some other source, or believe those two varieties are indigenous. Exposure must also come with an understanding of the socio-indices of the accents.

Age was a significant positive factor in answering the Awareness Task questions, possibly because older children may simply have more exposure to regional variation than younger children. Another possibility is that at least two of the Awareness Task questions involved making generalizations or using reason to say why a speaker with a novel accent sounds different than local speakers. Older children may be more proficient at making abstract statements or using this kind of logic.

Finally, another motivation for this task was to have information to use the general linear model to find factors that may influence discrimination ability. Using this statistical methodology, the effect the meta-linguistic knowledge demonstrated in this task has on ability to report hearing a regional accent will become apparent.

CHAPTER 4

ABX Discrimination Task

4.1 Introduction

This task asks whether five to seven year old children can discriminate between a familiar and unfamiliar regional accent. Establishing whether they have this ability is necessary, as previous research has suggested children at this age have difficulty with regional accent discrimination.

I expect that children can discriminate between regional accents, but that children who are aware of what an accent indicates about a speaker will perform better at discrimination than those with little awareness of regional accents. I use both a Parent Questionnaire (Chapter 2) and the Awareness Task (Chapter 3) described in the previous chapters, to measure children's levels of exposure and awareness of regional accent, and to look for correlations between those data and performance on this task.

This ABX discrimination task purposefully doesn't reference regional accents or geography in order to ensure that children are not biased in their responses by the framing of the question asked in the task. I want subjects to match speakers based on any criteria they find relevant. By allowing subjects this degree of freedom, I can test the measures of exposure and awareness for correlation with responses, to find evidence of influence they have on discrimination ability. I also can test if children find regional accent as salient as other kinds of variation in speakers they may perceive.

That children of this age can discriminate between two regional accents is the subject of some controversy. Extant research suggests that children go through a phase of language acquisition in which they are not attentive to regional variation in speech. Between three and five months of age, children are able to distinguish the local accent from another accent in their language (Butler et al., 2011; Egerova, 2010; Kitamura et al., 2006; Nazzi et al., 2000), but

cannot discriminate two unfamiliar accents of their language from one another (Butler et al., 2011). Several studies report that between 8 and 11 months of age, the ability to distinguish between local and non-local accents declines (Kitamura et al., 2006; Phan & Houston, 2008) (although see Butler et al. (2011) for counter-evidence), and this "deafness" appears to extend through at least 30 months of age (Phan & Houston, 2008), and possibly until after the sixth year of life (Floccia et al., 2009; Girard et al., 2008).

Other studies on children's sensitivity to the social meanings of accents lead me to believe it would be unlikely that five and six year old children cannot discriminate their local regional accent from another. In a series of experiments by Kinzler and colleagues (Kinzler, Corriveau, & Harris, 2011; Kinzler, Shutts, DeJesus, & Spelke, 2009), five year olds have been shown to have strong social preferences for and trust in speakers with the child's native accent, as compared with a non-native accent. This finding refers to Standard American English speakers in the United States; however in diglossic societies, such as South Africa, children of the same age report having social preferences for speakers of the language with the highest social status (Kinzler, Shutts, & Spelke, 2012). These studies provide evidence that not only are children sensitive to differences in accents, but attach social meaning to those accents, at the very least, in order to distinguish who is trustworthy. Similarly, other studies have shown five year olds make judgments about other speakers' characters, based on their speech varieties alone (Cremona & Bates, 1977; Day, 1980). Given that five year olds are aware of the social meaning of the language varieties heard in their environment, it would be odd if children didn't have similar reactions to regional accents of their native language, or were unable to discriminate them at all.

I suspect that the difficulty five and six year old children had discriminating accents in the two studies conducted by Floccia and colleagues (Floccia et al., 2009; Girard et al., 2008) is the result of how the task was framed. In these tasks, the children were asked to discriminate between utterances spoken in two regional accents, by sorting them into two groups based on where the speakers, represented by aliens, lived. One of the two varieties was the children's native variety, and the other a different regional variety from their country.

There are two potential problems with their stimulus materials. First, it assumes that children have the expectation that speakers (alien or otherwise) would speak differently based on where they live. Most five and six-year old children are generally not aware of the distribution

of regional accents, as shown in the previous chapter. If they detect the differences in accent, but do not know that the accent marks a speaker's home region, asking them to sort by where speakers live may appear to them like a guessing game. In the present discrimination task, I purposefully did not make any reference to accents or regions in the instructions the children heard at the beginning of experiments or in the individual trials. Because it is possible that some children are unaware of regional variation, or they interpret the distribution of regional varieties differently than adults, I don't want to confuse or influence their responses with adult interpretations of regional variation, making the assumption that their understanding of it is in line with mine.

Second, by using aliens, the link between a speaker's accent and home region is weakened. Children likely know that aliens are not real, or are not really from Earth. They likely also recognize the local accent used in the experiment as one from their hometown. However, the pairing of alien and familiar accent strongly suggests the speaker's home region is not important in this task, as the alien would clearly not be using the local regional accent. So although the tasks prompts the child to sort the speech samples by where the speaker is from, they may abandon any real-world knowledge they have of regional accents and choose a feature they might imagine to characterize alien speech.

To avoid this problem, I don't use any representation of the speakers in the stimuli materials, so that subjects are neither led nor discouraged from considering their knowledge of regional accents, and can discriminate the speakers based on any characteristic of the speech they choose. The results should show whether regional variation is a salient kind of variation to children of this age.

The present task uses an ABX design to test whether children can discriminate regional accents of their native language. The child hears a speaker (with either an accent local to their community or non-local accent, in this case from the Southern United States) say a single word. They then hear two additional tokens of the same word, one in the same regional accent, and one in the other regional accent. They must choose which of the second set of tokens best matches the first token heard. I believe that in this task, with the potential sources of confusion removed, that children will be able to reliably discriminate between two regional accents.

In addition to determining whether children can discriminate between regional accents, I am interested in whether awareness of regional accents or exposure to them positively influences

the ability to discriminate them. The reason for expecting both of these factors to enhance discrimination ability is based in Exemplar Theory (1997). This theory hypothesizes that listeners store memories, or traces, of speech they hear, and these traces will cluster with other similar traces to form exemplar clouds. Frequently heard tokens (of phonemes, lexical items, etc.) form dense clouds, which represent variants heard in the input. However, in addition to the acoustic properties, other information is thought to be associated with the traces, such as speaker identity, gender, race, and any other social properties characterizing the speaker. The social and the phonetic information associated with a particular variable is thought to be linked, so that perceiving one activates both linguistic and non-linguistic characterizations of the speaker simultaneously in the listener's mind.

Thus, if children have experience with different regional accents they will also have exemplar clouds for those accents, as well as some of the social information about speakers belonging to that category, such as where the speaker is from.

The children participating in this study should have a well-established category for the local accent, and know that the accent is associated with their hometown. This information could be used to help them find matching local speakers in the ABX Task. If they have experience with other regional accents, the question is whether it helps them to match the non-local accent heard in this task. I expect this might be the case, even for children who don't have experience with the specific non-local accent used in this study, but do have experience with other kinds of accents or social variation. Children may be able to extrapolate from this experience to help them identify the non-local speakers as not from their hometown.

Using a Mixed Effects Model, described in more detail starting on page 65, I also test whether prior exposure to regional variation, as assessed in the Parent Questionnaire, or awareness of regional accent, assessed in the Awareness Task, correlates with discrimination ability.

If I find a complete lack of correlation with any measures of exposure or awareness, it would likely indicate that the children are making matches based solely on the phonetics of the speakers' and that they are not utilizing any socio-indexical knowledge of regional accent that they may possess. This result would not mean that they don't possess regional accent categories, but just that they are not making use of them in this task. If this proves to be the case, it will

provide an interesting comparison between those subjects who are able to use socio-indexical information about speakers and those who do not.

Below are the questions and hypotheses for this study that pertain to this task:

Q1: Are five to seven year old children able to discriminate between a familiar and unfamiliar regional accent in their native language?

Q3: Is ability to discriminate between regional accents at all dependent on having direct experience with the regional accents in question?

Q4: Is ability to discriminate affected by children's overt awareness (i.e. ability to identify a regional accent and state what the regional accent indexes) of the accents in question?

The hypotheses responding to each of the above questions are as follows:

H1: Five to seven year olds will be able to discriminate between regional accents acoustically, when presented with a task that does not suggest any association with regions, accents, etc. This task should show that children are aware that regional variation is common across speakers, and that the differences between the stimuli speakers is not the result of some other kind of social variation.

H3: Given that Exemplar Theory has emphasized the role of storing and abstracting tokens of speech in the creation of categories, I would expect children who have direct experience with the non-local accent to be superior in its identification as compared to children without that experience. In particular, children who have family members from the South will have a higher rate of accuracy recognizing and grouping speakers with Southern accents.

All children in this study will have had intensive exposure to the local regional accent represented in the stimuli. However, I may find differences between children whose parents are not originally from the town where this research was conducted, or children whose parents are speakers of another variety of English (i.e. African-American English, L2 English, etc.) or another language entirely. These children, although they will have heard the local variety extensively in school, will have had most of their exposure prior to starting school in another variety. Therefore, I may see evidence that their exemplars of local speech may be defined differently than those children whose families speak the local variety, and this in turn may influence the correlations between experience and discrimination of the regional varieties.

H4: Given that there are effects of stereotypes of speakers with an accent in perception studies on adults, I expect that children who have an overt knowledge about accents also use this information when making decisions about who sounds similar. However, I expect relatively few children of this age to have an overt abstract knowledge about regional accents and their distribution, as they may lack

the geographical awareness to comprehend relative distances and location, making geographically based variation uninterpretable to them.

I expect to see less influence of this knowledge in the first task, where no reference to the socio-indexical value of the accent is made. I also expect children who are overtly aware of other kinds of variation, be it ethnic or L2 variation, likely don't transfer that knowledge to regional variation, since its sources are different, and regional variation still requires an understanding of geography to interpret.

4.2 Methods

4.2.1 Participants

Sixty-six children (35 female) age 61 to 77 months, average age 70.4 months, or 5;10, participated in this task. Sixty-one subjects completed the task. The subjects had all been enrolled in Kindergarten in public school near Philadelphia, U.S.A. for five months when the study was conducted. All but nine were born in the town where the experiment took place, although of those, all but one had lived over half their lives in the town.

The parents of the subjects came from a wide variety of backgrounds. The majority of the subjects completing this task (43/61) were Caucasian. This is the ethnic group associated with the local regional accent used in this study. There were 12 African-American subjects. All of the African-American students were observed to be speakers of African-American English (AAE) by the experimenter.

Eleven bilingual subjects also completed this task. The bilingual subjects all spoke English, but had different histories of language acquisition, so that identifying them as simultaneous or sequential bilinguals is difficult. I therefore have simply identified them as bilingual and did not attempt any further sub-categorization of this group.

Twenty-one of sixty subjects had at least one parent born outside of the region where the study took place. Of those, only seven were monolingual. I refer to this group of monolinguals with both parents from outside the region as the Outsiders; Insiders are subjects with at least one parent from the town where the child resides.

More detailed data on the subjects and the town in which they reside are given in Chapter 2.

4.2.2 Materials

A list of 30 stimuli words was created by identifying six vowel-quality differences between the Philadelphia and General Southern accents (Labov, Ash, & Boberg, 2006; Schneider, 2008). For each of the six categories, four to five common monosyllabic words that were easily illustrated with pictures and were not homophones with other stimulus words spoken in either accent were chosen as stimuli. The six vowel groups are characterized by the vowels in the following words, (words in block letters following Wells (1982)): *FACE, PRICE, GOAT, GOOSE, peel and tail*. In the *peel* and *tail* class, the critical difference is the vowel quality before /l/, seemingly reversed between the two accents in question, so that in Philadelphia "peel" is pronounced [pil] and in General Southern [pɪl], whereas for the word *hill* the pronunciation in Philadelphia is generally [hɪl] and in General Southern [hil]. The same is true for the *tail* class, but the two vowels that are interchanged are [e] and [ɛ]. For the *GOOSE* class of words, the /u/ is more fronted in Southern than in Philadelphia, and often preceded by the glide /j/, such that the pronunciation of *tune* becomes [tjun], for example. The *PRICE* class has a vowel that is pronounced as a diphthong in Philadelphia but as a monophthong in Southern [aɪ] vs [a:]. Finally the *GOAT* class of words has a vowel quality in Philadephia of [oʊ] that is considerably more fronted in Southern.

The five filler items were spoken by native speakers of Scots English. These items were unrelated to the target items, but were chosen for differing from both Southern and Philadelphia English, either in a vowel quality and/or production of a consonantal segment, particularly /r/, which was produced as [r] in these tokens. All words, both targets and fillers, were recorded in the carrier phrase "say ____ again" for uniformity of pitch. The stimuli were all sampled at 44 kHz. For a complete list of stimuli words, see Appendix 2.

Individual words (as opposed to sentences or longer clips of speech) are used as stimuli for several reasons. First, it allows children to make judgments on the accents based on vowel differences and not on sentential prosody or other supra-segmental phonetic differences between the accents that have not been not isolated in the creation of the stimuli. By limiting the stimuli to single words, more specific differences could be contrasted between the accents. Secondly, shorter clips of speech allow subjects to focus on the phonetic content of the stimuli, rather than the semantic content of the utterances. If the subjects are focusing on processing the sentences for meaning, they may not be able to devote attention to listening for phonetic differences in

accents. Also for this reason, they were also primed to expect hearing these words by being shown pictures and asked to name all of the items heard during the experiments.

Additionally, there was concern that in order to match sentence-length stimuli in an ABX task, they would have to retain a lot of information in their short-term memories, making this a test about short-term memory rather than accent discrimination. Based on a previous study of accent comprehension conducted with four and seven year olds using single-word stimuli (Nathan et al., 1998), in which four-year-old children were reported to hear differences in two regional accents in a word identification task, it seems likely that children in this study also would be able to hear that regional accent differences were present, even in an isolated word.

One could argue that using single words compromises the ecological validity of this study in some respects, since identification of a regional accent is often not based on hearing words in isolation, devoid of context. However, the main goal of this study is to show that children can perform discrimination when all of the complicating or contradictory external information about speakers was removed. Once it has been established that they can discriminate in the absence of extraneous information, future studies can establish whether they can also do so with longer and more complicated stimuli.

The stimuli for this task were created from recordings of six Caucasian male speakers, three for each regional accent, 25 to 35 years of age, and all lifelong residents of their respective hometowns. The local speakers were all from the same town as the children. The non-local speakers were all from the same town in northern Louisiana, and speakers of General Southern American English. The fillers were recorded by two Caucasian men from Scotland, aged 23 to 24, who had been living in the United States for two years at the time of recording.

The accents for the stimuli were chosen for this study because they are both well-described (Labov, 2001; Labov et al., 2006) and they are known to be highly salient to adult listeners (Clopper & Pisoni, 2004a; Labov, 1998; Preston, 1993). Speakers were all of the same gender, as previous research has found that even adults have a hard time overcoming gender differences when participating in dialect categorization tasks (Clopper, Conrey, & Pisoni, 2005) and I therefore expected children would experience similar difficulties.

The order of the trials was pseudo-randomized in one of four orders, ensuring that both Southern and Philadelphia accents were the "X" token 50% of the time, and that the matching token for a given word appeared equally as often in both the A and B positions across the entire

experiment. This was done in order to minimize effects of order, since it is possible that a short inter-stimulus interval (ISI) between the matching tokens might facilitate matching, whereas a longer ISI might make matching tokens harder.

The task was presented using Microsoft PowerPoint slides. For each sound clip presented in an ABX trial, a small icon representing the sound clip was placed under a heading labeling it as either A, B or ?, as shown below.

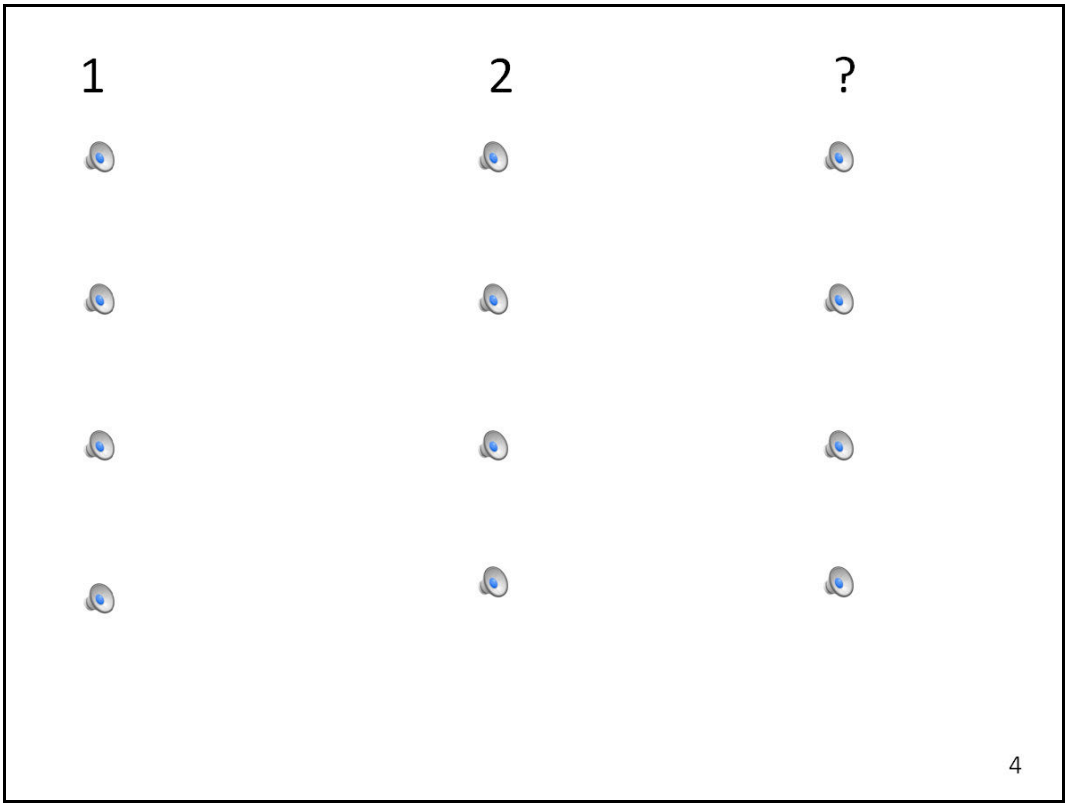

Figure 4.1 Example of Slide Used to Present Stimuli

Originally, the experiment was designed so that children would hear all 30 words twice, once in each order (AB & BA). However, in pilot testing, it was found that most Kindergarten-aged children do not have the attention span to complete 60 consecutive trials in one, much less two, tasks. Because the statistical design of this experiment depended on having within-subject data for comparison, the length of the experiments were shortened by half, and post hoc checks for effect of order were done to ensure that not having subjects complete both orders did not

adversely affect the outcome. Each of the four pseudo-randomized orders of the experiment was presented to one-quarter of the subjects, so that all orders were evenly distributed across subjects.

4.2.3 Procedure

Permission slips were sent home to parents several weeks prior to the experiment being conducted. Only those students whose parents had returned the consent forms participated in the study. The study was conducted during school hours, and the students taken individually from their classrooms in order to complete the experiments. The experiments were done in a quiet corner in a hallway neighboring the students' classrooms in their elementary school. This corner is often used by specialist and assistant teachers for individualized instruction, make-up work or assessment testing, so the subjects had previous experience receiving instruction or assessments at this location.

Each student was greeted and engaged in conversation before beginning the experiment, in order to make them feel comfortable and not have apprehensions about participating in the experiment. Once they had settled in their seat at the table, they were read the following script:

> *I am studying how children learn languages, and your parents and school have given me permission to ask you to play a game that will help me with my project. It has two parts, first a game, where you are going to hear people saying some words and you have to listen to them carefully and tell me which two people sound most alike. Would you be willing to do play the game?*
>
> *(wait for child's verbal assent)*
>
> *If you don't want to answer a question, or if you decide you don't want to finish playing the game, just tell me and we can stop. We can also take a break any time if you need to. OK?*
>
> *(wait for child's verbal assent)*
>
> *Are you ready to start?*
>
> *(wait for child's verbal assent)*

Before the experiment commenced, all children saw a PowerPoint presentation with pictures of each of the stimuli words in the experiment. They were asked to name the picture shown on each slide. This ensured that the children knew all of the words, but hadn't heard them

pronounced by any one except themselves immediately prior to completing the task. If the child incorrectly named a picture, or provided a synonym for the target word, they were asked to guess again until they said the target word. This was not common though, as the pictures had been pilot tested for recognizablilty by kindergarten-aged children beforehand.

The children then were given three warm-up trials, in which they heard words spoken by the same speakers who recorded the stimuli for the experiments. In the three warm-up trials, none of the vowel differences between the regional accents used in the test trials was heard. This was done in order to avoid biasing responses in the experiment, since for these warm-up trials, I gave the children feedback on whether they had correctly or incorrectly matched the tokens. No feedback was given during the actual experimental trials.

After the three warm-up trials, the children were asked if they were ready to start. The experimenter operated the computer, clicking on the icons to play the sound clips for the children, in order to minimize any effect that lack of prior exposure to computers might have had on the results. The children however, were given as long as needed to answer, although the clips were only played once; no child heard any trial more than one time. If he or she did not answer after about 15 to 20 seconds (although this was not explicitly timed), the child was asked if he or she wanted to continue. If they assented, the experiment continued and "no response" was recorded for that trial. There were only 28 trials across all 61 subjects in which no response was given.

During the experiment, if children made any comments or responded in any way to the speakers or the stimuli, this was recorded in a notebook for later analysis. It was also noted whether subjects were repeating words to make comparisons between their speech and the stimuli, or any other clues about how they were completing the task.

4.3 Description of Statistical Methodology

Because I am interested in factors that may affect children's abilities to discriminate between two regional accents, I collected large amounts of data on the subject's demographic and language backgrounds for use in statistical analyses of the discrimination task responses. The two instruments I used to collect these data were the Parent Questionnaire and the Awareness Task, both described in earlier chapters. From the responses, I was able to select independent variables that potentially affected the responses to the discrimination task. The

independent variables were selected through visual inspection of distributions of responses versus the variable in question, using tables and charts created from the raw data in MS Excel.

The statistical model used for analyzing the data is a multi-level logistic regression Mixed Effects Model. In this model, two kinds of effects on the responses are considered: random and fixed. The fixed effects are the independent variables I select; for example, age, gender, Awareness Score, etc. The random effects are those variables of which I have not taken measurements, but may exert influence on the responses provided. In this case, the item (the stimulus word heard in that trial) and the individual subject are random effects. This accounts for individual variation in the responses, as well as any difficulty particular words present across the subject group.

The specific Mixed Effects Model I used was a multi-level logistic regression model with binomial link function (in the library lme4, written by Bates et al., (2011) in the R Statistical Package (R Core Development Team, 2011). This model is designed to handle numerous trials with categorical responses, and therefore is well suited for this task, in which approximately1830 responses were recorded from 61 subjects.

4.4 Results

In this section, I first give the raw score results for the entire subject group, followed by the results of several sub-groups of subjects. Analysis of the subjects in sub-groups is motivated by a desire to examine the effects of language background on discrimination ability. Although I had also collected data for each subject about exposure to different dialects and regional accents in the Parent Questionnaire, these data were impressionistic, subject to the parents' interpretations and memories of what the child has experienced. Therefore, these data, while a useful guide, are potentially unreliable. For this reason, I also collected data on less subjective elements of the child's background. These data included the child and parent's places of birth and residence, whether a second language was spoken at home, and the child's race.

Using these measures, I could divide the subjects into groups, potentially reflecting their different language histories. The groups I chose for analysis are all paired with a contrasting group: Bilinguals and Monolinguals, African-Americans and Caucasians, and Insiders and Outsiders (children with both parents from outside the child's town of residence vs. children with

one or more parents from the region). In the following section, I explain my decision to analyze each of these sub-group pairings.

The differences in awareness between Monolinguals and Bilinguals might be found in amount of exposure to regional variation in English (by frequency) for the Bilinguals. I would suspect that another source of difference may stem from lack of exposure to cultural stereotypes, generalizations and depictions of regional variation in English, especially if the parents are not native speakers and do not propagate the cultural stereotypes of U.S. regional variation or have regular exposure to U.S. media depictions of regional variation in U.S. English. This could be tested by purposefully exposing bilinguals to depictions of U.S. regional variation, and then testing their awareness, although this was not done for the present study. In addition to the effects of a smaller amount of input in English, I also wanted to examine whether hearing another language or speakers with non-native accents in English heightened awareness of accents, either through direct experience contrasting them, or possibly from hearing commentary about accents and language varieties from adults.

In the Awareness Task, none of the Bilingual subjects could generalize the connection between geography and regional accent, and only 8% (as opposed to 51% of monolingual subjects) correctly identified the Southern speakers as non-local. Thirty-eight percent (as opposed to 77% of monolingual subjects) correctly identified the local speakers as being local. I therefore thought analyzing this sub-group separately was justified. Because the bilingual group varied in how they acquired English (simultaneously, sequentially, etc.), and they were small in number (n=11 in this task) I did not further attempt to sub-divide them for the purposes of the analyses.

African-Americans and Caucasians were contrasted because all of the African-American subjects spoke African-American English (AAE), and were therefore not speakers of the local regional accent, although they were all natives of that town. This presents an interesting case, as these subjects potentially have exposure to the local regional accent from living in the community, but likely do not speak it themselves. As a result, they may also hear commentary or have contrasted their variety with others in their environment, and like the bilinguals, have a heightened awareness of variation as a result.

The final sub-group pair looks at the effects of hearing regional accent differences regularly at home. Only monolingual subjects were included in these sub-groups, since

Bilinguals had been accounted for elsewhere. Children with one or more parents from the town where the study was conducted (Insiders) were contrasted with children having both parents from outside of the town (Outsiders). This was done to take into account experience hearing different regional varieties on a regular basis as well as commentary on regional differences from non-local family members and Outsider parents.

The three sets of sub-groups therefore represent children with different experiences hearing contrasts in the varieties spoken by the local majority and those spoken in their homes: Bilinguals with non-native accents, African-Americans with ethnolects, and Outsiders with regional accents. There was no overlap in these three groups (i.e. all of the African-Americans were Insiders and monolingual, none of the Outsiders were bilingual or African-American). Although I cannot exactly measure each individual subject's experience with hearing contrasting accents and commentary on them, I designed the sub-groupings to be rough approximations of that experience. The intent is to show how this experience might cause differences in the ability of the contrasting sub-groups to extract socio-indexical information from accents, or in their interpretation of regional accents.

Below I present the results of the analyses for the entire subject group, followed by the same pairings of sub-groups addressed above. Each is followed by a short explanation and comparison of the sub-groups.

Following a review of the results of the ABX Discrimination Task, the statistical model will be presented, followed by its results for the entire group and sub-group pairs. I also include an item analysis, looking at the effects of inter-stimulus interval (ISI), vowel contrasts and speaker pairings had any effect on performance.

Note that in all of the results reported, only the responses to 25 test trials are reported, as the five filler tokens were not included in the analyses.

4.4.1 Entire Subject Group

Sixty-six subjects participated in this study, although only 61 subjects completed this task. Five were later removed from the analysis for the following reasons: irregularities in testing (1), failure to pay attention during the task (2), being later identified as needing speech therapy (1), and missing the parent questionnaire (1) making impossible to include that student in

the statistical analyses. The average number correct for the entire subject group was 16.18/25, or 64%. The high score was 22/25; the low score was 7/25.

A chart showing the distribution of scores is given below:

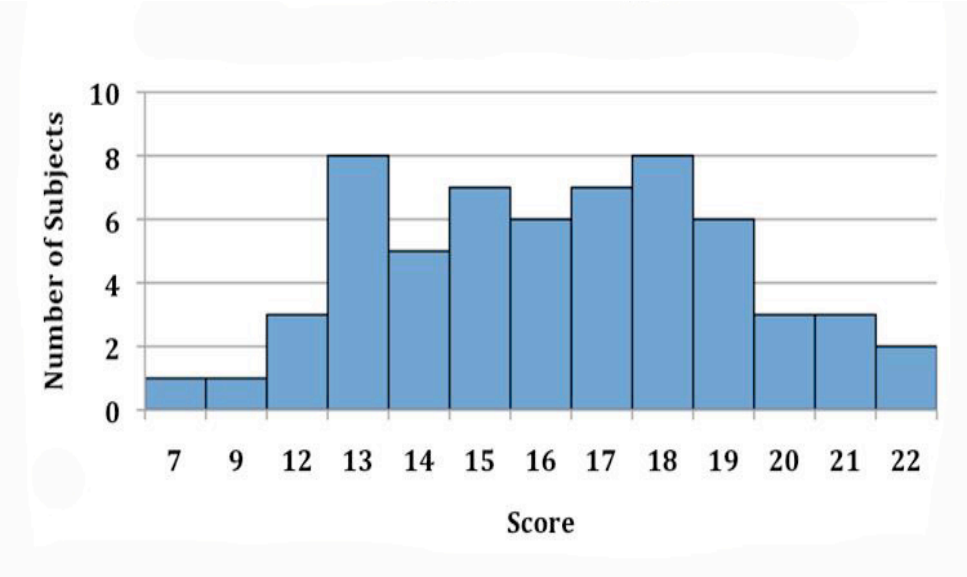

Figure 4.2 Histogram of Scores for ABX Task

I had initially created Awareness and Exposure index scores, based on the responses to the Awareness Task and the Parent Questionnaire. I had intended to capture relative levels of these two factors, and compare their respective effects on discrimination ability. As it turned out, neither index score showed any sign of trending with performance on the discrimination task. I therefore turned to examining the individual measures of awareness (the five questions on the Awareness Task) and exposure (items from the Parent Questionnaire concerning how often children heard other accents).

Surprisingly, none of the measures of exposure, assessed from the Parent Questionnaire, warranted inclusion in the model based on the initial triage of fixed effects. Likewise, no demographic factors, such as age or gender showed any correlation with the distribution of responses. The only factors that showed any correspondence to the distribution of responses were two of the Awareness Task questions. These were analyzed as fixed effects in the model.

The two questions analyzed were Q3 and Q5 from the Awareness Task. Q3 asked children to specifically state where the local speaker was from; any indication that the child gave that the speaker was from the same town as the child was counted as correct. Q5 asked the children to generalize and state why the two sets of speakers heard in the experiment sounded different; any answer in which the children identified the speakers as being from different places was counted as correct. Seventy-one percent of subjects got Q3 correct, whereas only 34% correctly answered Q5.

In this analysis, only Q5 of the Awareness Task (the question asking children to identify why the two sets of speakers heard in the task spoke differently, to see if they could generalize the differences instead of just identifying whether the speakers were from their hometown) was significant. However, the influence was in the opposite direction as expected: knowing that the main difference between the two sets of speakers was regional accent made correct matches in the trials *less* likely.

However, a second factor nears significance for this group: correctly identifying the local speakers as being from the child's hometown (Q3) *positively* influences a correct match on a trial. This correlation is in the expected direction, as the identification of two speakers as local may have provided a clue as to which speakers were the matching pair.

Table 4.1 Whole Subject Group

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.2244	0.1268	1.769	0.076839 .
Q5	-0.2685	0.1284	-2.091	0.036556 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.4.2 Bilinguals and Monolinguals

This group was included to compare the effects of having English compose only part of the child's regular input, whether those effects are deleterious due to reduced exposure to English, or enhancing, as a bilingual child is possibly more aware of language varieties and variation than a monolingual child. Bilingual children possibly have more experience traveling outside of their home region and interacting with speakers of another language, perhaps making

them generally more aware of language variation, although a different kind of language variation than what is addressed in this study.

The bilinguals (n=11) averaged 15/25 (60%) correct on this task. The monolinguals (n=50) averaged 19.8/25 (79%) correct. The difference in average scores between the two sub-groups was significant: $t=-3.5047, p= 0.00438$ in a two-tailed t-test. The averages for each group were also tested against the average of the whole subject group. For the Bilinguals vs. the whole group, $t= -0.8612, p=0.4059$ in a two-tailed t-test. For the Monolinguals vs. the entire group, $t= 6.4191, p=3.977 \times 10^{-9}$ in a two-tailed t-test. The Bilinguals were not significantly different from the group as a whole, whereas the Monolinguals were significantly better than the entire subject group.

Because the numbers in some sub-groups were small, and tests for parametricity were inconclusive, I also used the Wilcoxon Rank Sum Test to test differences in group means. The results of this test converge with those of the two-tailed t-test, but I report them here as well. In the Wilcoxon Rank Sum Test, the means of the two groups were significantly different ($W=95, p=.00082$). For Bilinguals vs. the entire subject group $W=398.5, p=0.2772$; a non-significant difference. Monolinguals were significantly better than the entire subject group ($W=569.5, p=3.71 \times 10^{-8}$).

4.4.2.1 Monolinguals

The same pattern seen in the subject group as a whole repeats itself here, but with the correlation between Q3 (identifying the local speaker) and correctly discriminating between the accents reaching significance. I suspect that because Monolinguals make up such a large portion of the entire subject group (50/61), that they also drove these trends in the larger group.

Table 4.2 Monolinguals

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.2754	0.1380	1.996	0.04590 *
Q5	-0.3388	0.1270	-2.667	0.00766 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.4.2.2 Bilinguals

There was some difficulty in finding any factors that had potential for affecting responses on this task. For the Bilinguals, Q5 could not be tested in the statistical model because there was no variation in the answers provided by members of this sub-group; none of the Bilinguals answered this question correctly.

No other factors appeared to correlate with performance in the initial triage of the data and therefore were not analyzed in this model.

Table 4.3 Bilinguals

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	-0.3682	0.4685	-0.786	0.4319
Q4	0.5062	0.7319	0.692	0.4892

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.4.3 African-Americans and Caucasians

The African-American subjects (n=12) averaged 16.09/25 (64%) correct on this task, and the Caucasians (n=38) 16/25 (64%). This was a non-significant difference between the two sub-groups, in both a two-tailed t-test (*t= -0.3315, p=0.7455*) and the Wilcoxon Rank Sum Test (*W=233, p=0.948*). This is also not significantly different from the score of the whole subject group. For African-Americans vs. the entire subject group: *t=-0.0818, p=0.936* in a two-tailed t-test; *W= 323.5, p=0.924* in a Wilcoxon Rank Sum Test. Caucasians vs. the entire group: *t= 0.5016, p= 0.6171* in a two-tailed t-test; *W=1246, p=0.770* in a Wilcoxon Rank Sum Test. Again, the t-test and Wilcoxon Rank Sum Test results converge. Neither African-Americans nor Caucasians were significantly different from the group as a whole.

4.4.3.1 Caucasians

For this large sub-group, the unexpected significant, negative correlation between correctly answering Q5 and making correct matches appears, as with the Monolinguals and the subject group as a whole. Explanations for this correlation are presented in the discussion of the results.

Table 4.4 Caucasians

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.04887	0.14348	0.341	0.7334
Q5	-0.30561	0.13480	-2.267	0.0234 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.4.3.2 African-Americans

For African-Americans, there is a positive effect of identifying the local accent as being from the child's hometown (Q3) on making correct matches. It could be that this particular sub-group was able to identify local speakers and make matches based on that information, rather than depend on direct comparisons of the tokens heard in each trial. It is unclear however why they don't show the same level of negative correlation between Q5 and correct matches as the entire subject group does. It could simply be a matter of the small number of subjects in the African-American sub-group, and with additional subjects, they would show the same significant, negative correlation as the other sub-groups.

Table 4.5 African-Americans

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.6076	0.3049	1.993	0.0463 *
Q5	-0.3856	0.3055	-1.262	0.2069

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.4.4 Insiders and Outsiders

The Insiders are subjects with at least one parent who was born in the town where the study was conducted. Outsiders are subject with both parents from outside of the town. Children with only one parent raising them were grouped depending on where that parent was from. Only Monolingual subjects were included in the Insider/Outsider comparison. These sub-groups are compared to test whether regular exposure to other regional varieties, both at home

and possibly through travel and interaction with other relatives speaking other varieties of English, affects discrimination.

The Insiders (n=40) averaged 16.5/25 (66%) correct. The Outsiders (n=12) averaged 16/25 (64%). Again, these were not significantly different from one another (*t= 0.5074, p=0.6138* in a two-tailed t-test), or from the whole subject group (for Insiders: *t= 0.3173, p=0.7523* in a two-tailed t-test; for Outsiders, *t=-0.2692, p=0.7887* in a two-tailed t-test).

Using the Wilcoxon Rank Sum Test, Insiders and Outsiders were not significantly different (*W=462.5, p=0.805*). Neither Insiders nor Outsiders were significantly different from the subject group as a whole (Insiders: *W=793, p=0.879*; Outsiders: *W=1007, p=0.894*). Again, these match the results of the two-tailed t-test.

4.4.4.1 Insiders

The Insiders comprised a majority of the subjects in this task, as well as of the Monolingual sub-group, so it is unsurprising to see the same pattern of correlations for this sub-group as with those two sub-groups.

Table 4.6 Insiders

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.3491	0.1624	2.150	0.0315 *
Q5	-0.3352	0.1482	-2.261	0.0237 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4.7 Outsiders

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.4319	0.2479	1.742	0.0815 .
Q5	-0.1850	0.2486	-0.744	0.4567

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Outsiders included only the monolingual Outsiders, in order to compare the effects of hearing different regional accents at home, and not non-native accents or second languages, which are accounted for in the bilingual group. All of the Outsiders answered Q3 correctly, and because the model needs variation in the answers in order to analyze the data, Q3 could not be included as a fixed effect in this analysis, leaving only Q5 as a fixed effect in the model. It appears that Q5 had no influence on the Outsiders' matching ability, although this may also be a result of the relatively small sample.

4.5 Item Analyses

4.5.1 Interstimulus Interval

As this was an ABX design, I tested for effects that inter-stimulus interval (ISI) had on children's potential to correctly identify the matching tokens. Using the same statistical model described above, and using the ISI as the fixed effect, I found that the longer ISI facilitated the correct match. Although the correlation was just shy of reaching significance, the trend indicates that if the child heard the matching token after the non-matching one, the likelihood was higher that the correct match was made in that trial than if the matching tokens were heard in succession.

Table 4.8 Effect of ISI

	Estimate	Std Error	Z Value	P Value
Long ISI	0.2092	0.1075	1.945	0.0517 .

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.5.2 Effect of Matching Speaker Pairings

In this analysis, I examine whether some speaker pairings were more difficult to match than others. It appears that only one pairing, two of the Southern-accented speakers, presented a significant difficulty to subjects. Both possible orders of speaker pairings are included, as each subject heard only one order of the pairings.

Table 4.9 Effect of Speaker Pairings

	Estimate	Std Error	Z Value	P Value
SpeakersBJ	-0.13992	0.46637	-0.300	0.76416
SpeakersCB	-0.33337	0.46134	-0.723	0.46991
SpeakersCJ	-0.09015	0.33167	-0.272	0.78578
SpeakersJB	0.67033	0.39112	1.714	0.08655 .
SpeakersJC	-0.17445	0.36907	0.473	0.63644
SpeakersFQ	-0.47314	0.45950	-1.030	0.30315
SpeakersFU	-0.31452	0.32932	-0.955	0.33955
SpeakersQU	-0.45378	0.36601	-1.240	0.21504
SpeakersUF	-0.70323	0.28559	-2.462	0.01380 *
SpeakersUQ	-0.04031	0.46759	-0.086	0.93130

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.5.3 Relative Difficulty of Southern vs. Local Accent

This analysis examined whether one accent was consistently harder to match than the other. It appears that Southern speakers were more difficult to match compared to the local speakers. The rate at which the children correctly matched the Southern speakers vs. the local speakers was 67% correct for the local accent and 56% correct for the Southern accent.

It is not surprising that Southern was relatively more difficult, in that it may have been the first time some subjects had heard this accent. According to the Parent Questionnaire, only 7/66 subjects were reported to have had any prior exposure to a Southern accent, and none had intense or prolonged exposure to a Southern accent.

Table 4.10 Comparative Difficulty of Accent

Accent	Estimate	Std Error	Z Value	P Value
Southern	-0.5170	0.1910	-2.707	0.00678 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Because there was a difference in difficulty matching the two speakers, I divided the trials up by accent to be matched, and ran the Mixed Effects Model to see if the fixed effects found in the subject analyses were specific to either of the accents heard in the discrimination task. This indeed was the case. Recognizing the local speaker as local in Q3 of the Awareness Task positively influenced discrimination in the trials where the matching speakers were local. However, knowing that the two sets of speakers in this experiment were from different regions (Q5) negatively affected discrimination of the non-local speakers. I offer further explanation of this result in the discussion section below.

Table 4.11 Trials in Which Matching Speakers were Non-local

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	-0.05503	0.19272	-0.286	0.77522
Q5	-0.47933	0.19496	-2.459	0.01395 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4.12 Trials in Which Matching Speakers were Local

Question	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.5688	0.1823	3.120	0.00181 **
Q5	-0.0240	0.1894	-0.127	0.89915

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.5.4 Effect of Vowel Class

Here I tested whether certain vowel types were more difficult than others for children to discriminate in this task, as compared to filler items. It appears that all of the test stimuli were more difficult than the fillers, although the least difficult of all were the PRICE type words. The most difficult were the GOOSE type words.

Table 4.13 Effect of Vowel Class

	Estimate	Std Error	Z Value	P Value
DAY	-1.3600	0.2851	-4.771	1.84e-06 ***
PRICE	-0.5649	0.2783	-2.030	0.04241 *
GOAT	-1.2012	0.2863	-4.196	2.72e-05 ***
Peel	-1.4260	0.2850	-5.003	5.64e-07 ***
Tail	-0.8438	0.2893	-2.917	0.00354 **
GOOSE	-1.5591	0.2851	-5.468	4.55e-08 ***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.6 Discussion

In this task, kindergarten-aged children are asked to match speakers based on regional accents, but without being told that regional accent or region of the speaker was related to the differences between the speakers. The results of this task show that children of this age are able to reliably discriminate between regional accents. The entire group averaged 64% correct, and when considering the Monolingual subjects alone, the average was 79% correct. None of the sub-groups were significantly different from the entire subject group, and only in one case were the two members of the sub-group pair significantly different (Monolinguals and Bilinguals).

The Bilinguals were significantly worse than the Monolinguals at this task, scoring only 60% correct. Note that they perform better than chance, even if they are not as accurate as the Monolinguals. The difference between Monolinguals and Bilinguals is perhaps unsurprising, given that a bilingual speaker may have relatively less exposure to American English and its regional varieties if a large percentage of the input is in another language. I did not try to distinguish for this group whether the result was due to non-native-like vowel categories, which themselves might distort perception of vowel differences in the L2 or if they had less awareness of U.S. regional variation and therefore could not draw on the socio-indices of the regional accent to assist in making discrimination judgments, as many Monolinguals did.

The second question addressed in this task is whether any factors that improve discrimination ability can be identified. No exposure factors, as assessed via the Parent Questionnaire, appear to exert any influence. It could be that the measures I chose to examine do

not affect discrimination ability, or that the parents' responses don't reflect actual amounts of exposure. However, I also looked at broader groupings that more generally represented exposure to different accents, and there was no difference in the rate at which children belonging to these sub-groups discriminated between the two regional accents either, Bilinguals notwithstanding.

Given the lack of significant differences in the accuracy of the sub-groups, it also appears that exposure to other native English varieties as the main source of language input doesn't affect discrimination ability. This was even the case for children with exposure to other regional accents. Unfortunately, the number of subjects with even a single day of prior exposure to Southern accents was too small to analyze separately, so I couldn't compare those subjects to other sub-groups to see the effects of exposure to the non-local accent used in this study. This would be a fascinating comparison to make in future work.

Awareness of accents, represented by an index score created from the responses to the five questions of the Awareness Task, also didn't correlate with discrimination ability. However, when the Awareness Task questions were examined individually, two had a significant effect on discrimination of regional accents in this task. I had expected that the correlations between measures of awareness and discrimination to be generally positive: more of the former results in better performance on the latter. But the two Awareness Task questions (or the knowledge they represent) used in the model influence discrimination in opposite directions, one positive and one negative.

Q3 of the Awareness Task asked children to state where the local speaker was from, to see if they were aware that a regional accent was spoken in their hometown. This question tests whether children can interpret the basic social index of a regional accent: where a speaker is from. Children who answered this question correctly were more likely to correctly discriminate between the two regional accents in the discrimination task. This was the hypothesized result; a child who knew there was an accent associated with his hometown might have at least a binary understanding of regional accents (local vs. non-local), and draw on this information to assist with discrimination.

Q5 asked children to make a general statement about why the two sets of speakers sounded different, in an attempt to assess if children were aware of how regional accents were distributed. This question doesn't directly address socio-indexical information, but asks them to generalize about it, and was included to see if a general awareness of accents also affected

discrimination. Answering Q5 correctly decreased the likelihood that children made the correct match in discrimination task trials.

I found similar patterns of correlations in the sub-groups' results. The Monolinguals and Insiders both show the same correlations as the whole subject group. The African-Americans have only the positive Q3 correlation and Caucasians the negative correlation with Q5.

It seems counter-intuitive that two measures of accent awareness would influence discrimination in opposite directions. However, if it can be shown that the correlations between these two measures and performance on discrimination trials are indicative of strategies children use to discriminate between familiar and unfamiliar accents, as opposed to representing general awareness of regional accents, a possible explanation emerges.

Q3 asks children to access the socio-indexical information about the local speaker: identifying that person's place of residence. The fact that this one question has a positive influence on correct discrimination may indicate that children are using the socio-indexical information provided by the local regional accent as a short cut to discrimination. If they can identify two local speakers, they can match them based on the realization those speakers are from the same place as the child.

Note that this same correlation does not obtain for Q4, in which children must identify the Southern speakers as non-local. The absence of an effect for Q4 indicates that children are only able to use socio-indexical information for well-established exemplar categories. Because few of the subjects had any familiarity with Southern U.S. accents, they may have lacked the category and concomitant socio-indexical information to help them match the speakers.

This leaves the negative effect of Q5 on correct discrimination to explain. The difference between Q3 and Q5 however is the kind of information addressed. Q5 asks children to make an abstract statement about how regional accents are distributed; it asks them to associate regional accents with geographical location. It isn't asking them to access information about specific speakers; it is asking them to extrapolate that knowledge in order to make a general statement. This may seem like a fine distinction, but the kind of experience needed to acquire these two skills is quite different.

In the case of associating a regional accent with the hometown, this knowledge could be gained via experience of hearing the local accent repeatedly in the hometown. To answer Q5 correctly, children were either told that people from different places speak differently, or they

have enough exposure to people from different regions of the U.S. to have figured out the connection on their own. Although they may know regional variation exists, they may not know what specific phonetic differences characterize another regional accent.

Based on parent responses about children's interaction with Southern U.S. speakers or media exemplifying Southern U.S. accents, few, if any, subjects in this study would have an exemplar category for a Southern U.S. accent, and therefore no socio-indexical information associated with that accent. This could mean that because the Southern accent didn't match any previously experienced variety, and the children had no information about the speakers coming from the same region, some subjects were unwilling to match the non-local speakers in this task.

Other subjects, aware that regional accents exist, but also lacking an exemplar category of Southern speech, may have tried to guess which features to match for the novel accent stimuli. This conjecture may have led to the negative correlation found for Q5 in this task. The subjects who correctly answered Q5 may have focused on the wrong features as characterizing the unfamiliar, non-local accent, leading to a lower rate of correct matches, as compared with the local accent stimuli. In other words, the combination of awareness of regional accent distribution and lack of experience hearing the particular accent in question may have complicated discrimination for the 34% of subjects who answered Q5 correctly.

Correlation, of course, does not imply causation, and the results of the statistical analysis do not prove that this is what the subjects were doing in this task. However, I use the item analyses to find corroborating evidence to strengthen the argument that the correlations found represent strategies children were using in the discrimination task.

In the item analysis, described in section 4.5.3, I tested whether the accent to be matched had an effect. Matching speakers with the Southern accent, the non-local accent, was shown to be more difficult than matching speakers with the local accent. This led me to conduct a second analysis, in which I compared whether Q3 and Q5 correlate trials on a specific accent, i.e. if the positive Q3 correlations were only found for trials in which the local accent is being matched. Indeed, there was a difference in correlations by accent: for those trials in which the matching pair were Southern speakers (non-local), Q5 negatively correlated with correct discrimination. In trials where the matching speakers were local, Q3 positively correlated with correct discrimination.

Based on this division by accent, there is evidence that children may be using different strategies depending on the accent to be matched in the ABX trial. It seems as if they are drawing on their detailed knowledge of the local accent and its speakers in order to discriminate in trials where the local speaker is to be matched. In trials where it is the non-local speaker is to be matched, Q5 negatively correlates with correct matches. Subjects who are aware that the unfamiliar sounding speakers may be from somewhere else (i.e. those that answered Q5 correctly) may try to guess what features to match, possibly from their representations with other regional accents. This tactic proves to be ineffective. Subjects who didn't answer Q5 correctly may not be trying to draw on past experience to identify the accent and perhaps use some other technique for making the match.

This dichotomous use of strategies could be indicative of what children's exemplar representations of regional accents look like at this age. First, they seem to have a detailed representation of the local accent's phonetic characteristics, and have some socio-indexical information linked to that accent, at the very least information about where the speaker is from. It also appears they are able to use this socio-indexical information to facilitate discrimination, in effect using the socio-indexical information to provide a short-cut for identifying similar accents.

The results of the statistical analyses also show that although some children can explicitly state why two speakers with different regional accents don't sound alike (Q5), this information does not help them with discriminating regional accents. It in fact makes finding the two matching speakers in the ABX task harder. I contend that Q5 represents another kind of knowledge, which is not based on experience with the accents, but like a stereotype, provides a general kind of information about the accent. Although the listeners try to draw on this general information about the accent, it appears to not help them to correctly find similarities between speakers with an unfamiliar accent. Because it is not linked to an accurate phonological representation of the accent, it can't be used as a short-cut to identify matching speakers in the same way that socio-indexical information is.

One possible explanation in the difference between the effects of Q3 and Q5 is the kind of information these two questions represent. Q3 requires only that the children recognize the local speaker, and that they have associated the local speech with the correct socio-indexical information about the speakers. They are able to access the socio-indexical information about local speech to state which speakers are local when asked, but they can't use it to reason that the

speakers with unfamiliar accents are non-local. Similarly, answering Q5 would be easy if they could reason about what the regional accents index. But it seems even the children who did answer Q5 correctly were not using reason to answer the question: not all of the children who answered Q3 correctly also answered Q5 correctly, and vice versa.

It is also an important finding that children seem to use socio-indexical information about speakers to complete the matching task whenever possible, and as seen above, even when it degrades performance. Perhaps it is surprising that they show such a strong tendency for using social information in the speech signal, but I believe that this underscores the importance of social knowledge in speech perception - and the possibility that social knowledge direct listener's perception, independent of the acoustic information in the speech signal.

4.6.1 Children's Exemplar Representations of Regional Accent

The results of the discrimination task can help outline what the exemplar categories might look like for children this age. It seems that children are associating information about speakers' region of residence with regional accents, but only for those accents with which they are familiar. This would follow from the proposed method of how exemplar clouds are constructed: through the accumulation of tokens, tagged with information about the speaker. For a regional accent with which they have little experience, or which they have heard, but have no socio-indexical information about the speakers, there is no recognition of what that speaker's accent represents. This strongly suggests that it is experience hearing and associating a regional accent to specific set of local speakers that forms a category.

Based on the lack of correlations between discrimination and Q4 (identifying that the non-local speaker is not from their hometown) there is no evidence of a general "non-local" category that is activated whenever a non-local accent is heard. This again suggests that categories are accent-specific, and require exposure to a specific regional accent to create and associate with socio-indexical information.

The facilitory effect of Q3 and lack of effect of Q4 (identifying the non-local speakers as non-local) on discrimination also suggests that children don't need to have a contrasting, non-local exemplar category to have formed a category for local accents. That is to say, children can interpret a regional variety as being associated with their hometown even if they don't know that other regional accents are associated with other towns and regions, or even if they have no

experience with other regional varieties. They may have only one exemplar category for regional accent at this stage, and that is for the local variety.

Finally, although children may be aware that regional variation exists, and they can hear it, they only construct categories for varieties they have experienced. Experience seems to be the most important factor in identifying regional accents at this age. Children do not appear able to reason about accents, or focus on their phonetic details to find similar sounding accents. It is particularly noteworthy that at a relatively early age, children have gone from focusing on phonetics to listening for social information about speakers in order to identify linguistic varieties, and in its absence have difficulty finding similar sounding speakers. Children appear to use social categories to structure their linguistic world in much the same way adults do.

4.7 Conclusion

This task established that children can discriminate between a familiar and unfamiliar regional accent of their native language. No correlations were found between any demographic or exposure factors and discrimination performance, nor were the accuracies on the discrimination tasks significantly different between most of the sub-groups. The only exception to this was the bilingual children; they were significantly less accurate at discrimination than monolingual subjects and the group as a whole. Those discrepancies may be attributed to other factors that were not controlled for in this study. Further work is needed to determine the source of the difference.

Additionally, I tested whether awareness of regional accents improved performance on discrimination. Although awareness in general did not improve overall performance, particular kinds of knowledge demonstrated on the Awareness Task did improve performance on some trials in this ABX Task. Children who could identify the local speakers as being from their hometown did better matching the local speakers in this task. This awareness did not generalize to help match the non-local speakers. This suggests that knowledge about regional accents is very specific, based on experience hearing a particular accent and associating it with speakers from a particular location.

It also appears, based on the fact that the matching accuracy rate was significantly worse for the non-local accent, that recognition of the speakers' region of origin improves performance. It could be that recognizing this social information about speakers provides a short cut in

matching, relieving children of the need to compare phonetic features, and allowing them to compare similarity on another basis.

Interestingly, general knowledge about regional accents degrades performance in matching unfamiliar speakers. Although children will attempt to utilize this knowledge when discriminating between speakers with an unfamiliar accent, it appears to complicate finding the correct match.

Finally, there is no evidence that exposure to other kinds of variation improves performance. Children don't seem to use any knowledge about other kinds of accents to assist in discrimination of regional accents. In general, it appears that children don't reason about novel accents based on information they have about familiar accents. There is no evidence that they will extrapolate from what is known to interpret the unknown.

CHAPTER 5

Similarity Judgment Task: Can Children use Socio-indices to Identify Local Speakers?

5.1. Introduction

The previous task presented the results of an ABX discrimination task, which established that children aged five to seven can discriminate between two regional accents of their native language. The ABX task was designed to avoid any reference to accents or speaker locations, in order to test whether regional accent was salient to children and whether they would match speakers by accent, even when not explicitly directed to do so.

However, the very nature of regional accent is that it provides information about the speaker. It does not exist simply as an acoustic difference in speech. It is not known, however, whether young children have acquired any of the social meanings that an accent carries, also known as indices (Silverstein, 2003b). This task will examine whether children aged five to seven recognize that a local accent marks speakers as being from their town of current residence. For brevity, I will refer to this as the 'hometown.'

A regional accent first and foremost indicates the speaker's place of birth. However, regional accents may also be interpreted as signaling other personal qualities of the speaker, such as intelligence, socio-economic status, etc. This information about speakers, both subjective and objective, is connected with the set of phonetic features characterizing that accent, and they are inextricably linked for listeners familiar with those indices.

When an accent is perceived, the social information about the speaker, both of the first and second index variety, is activated, affecting perception of the speaker and expectations about that person's speech (Campbell-Kibbler, 2007; Staum-Casasanto, 2009). The reverse is also true: the presence of social information about a speaker has been shown to trigger perception of vowels associated with a particular accent, even when those vowel sounds were not actually heard (Rubin, 19992; Niedzielski, 1999; Hay and Warren, 2010). Johnson (2006) shows how

this is likely the result of exemplar representations existing for both linguistic categories (whether they be phonetic, lexical, grammatical, etc.) and social categories. Activation of a social category during perception biases how the listener experiences the speech signal, which in turn biases how they judge the speaker's social qualities.

It is not well known how and when children come to establish adult-like social and linguistic exemplar categories for different kinds of social variation. Foulkes and Docherty (2006) suggest that because exemplar categories are formed from experiencing phonological variants and interacting with individuals representing specific social categories, that associating social indices with linguistic tokens may take an extended amount of time, especially for social categories that are not visually transparent. If this is true, then regional variation may be one of the last categories children in the United States establish for two reasons. First, region of residence is difficult to establish based on appearance of the speaker. Second, most children will not have much experience traveling or interacting with people from outside their home region.

In the first task of this study, reported on in the previous chapter, children could discriminate between two regional accents in the absence of any other information about the speakers. It is possible that they could have been basing their judgments on the acoustic similarity of the tokens in each trial. However, the correlations that were found suggest that some subjects were using non-linguistic information about the speakers to complete the task. The correlations between successfully matching speakers and the knowledge that one accent was local suggests that they used identification of the local accent as the means by which they matched speakers. This suggests that children have a linguistic category for local accent and it is linked to socio-indexical information about the speakers. Hearing the local accent activated that information about the speakers, facilitating discrimination. The same effect was not found for the non-local speakers.

The most surprising thing about the correlations found in the ABX task was that children were not prompted to make these associations. In fact, every effort was made to avoid suggestion of regional accent. Yet the socio-indexical information about speakers was still found to exert an influence on responses, as Exemplar Theory might predict.

In the present task, I prompt children to access the socio-indexical information they may have associated with regional accents. In this task, children hear two clips, one of the local and one of the non-local speaker, and must judge which person sounds most similar to the child. The

task allows me to observe whether children will use the social indices of regional accent to identify speakers who are local to their hometown, by asking whether they will categorically choose local speakers as sounding most similar. If they overwhelmingly choose the local speaker as most similar sounding, despite the mismatches in age, gender, etc. between themselves and the speakers, this will indicate that they can draw on the social meaning of the local regional accent.

As with the previous task, the correlations between the Awareness Task questions and the Parent Questionnaire will be examined statistically using a Mixed Effects linear regression model. If there are correlations between awareness, exposure and the responses on this task, these would provide evidence that children are using knowledge about the links between regional accents and geography to make similarity judgments. In the following paragraphs, I outline what the results might look like if children are relying on methods other than using their knowledge of regional accent indices to make their similarity judgments.

It is possible that children could be conducting phonetic comparisons between their own speech and the tokens in the experiment. If this is the case, few correlations between exposure and awareness and their responses in this task should be found.

Another possibility is that they will choose the most familiar accent as sounding most similar. This could possibly lead to a categorical pattern of judging the local speakers most similar as well. However, if this is the case, I should find few correlations between their responses and the independent variables from the Awareness Task and the Parent Questionnaire, as this knowledge would not affect their judgments.

The correlations between judgments made about speaker similarities and awareness of what accent represents should show whether children link regional accent with a particular location. Additionally, it will show that children have developed the expectation that people from their community speak similarly. However, what I won't know is how narrowly defined subjects' notion of the regional variety is. To assess that, I must look at subjects with input from more than one regional variety, or who are exposed to other non-majority varieties regularly, and examine their pattern of responses and correlating factors with the results of this similarity judgment task. These children might not have well-defined representations of local speech, because they are aware that multiple varieties are heard in their hometown, and therefore think it possible that a novel variety might also be from their hometown. Alternately, they could have

very specific representations of which varieties are heard locally, and be just as unwilling as children whose input is only in a single regional accent or variety to recognize the Southern speakers as local.

5.2 Background

As shown in earlier chapters, people's ability to characterize their own accents is highly subjective, and heavily dependent on prior experiences with regional accents. The subjects of this study, although only between five and seven years old, have varying amounts of experience with regional and other kinds of sociolinguistic variation. The stimuli used to represent the local accent do not account for the clines of regional accent, much less the different ethnolects found in that community. It is therefore the case that some subjects won't hear stimuli that match their native speech variety. This was intentional; the similarity choices made by linguistic minority subjects are potentially very informative on how they interpret sociolinguistic variation in their environment.

The criticism can be made that in a forced-choice task, children from the linguistic minority groups have no opportunity to express dissimilarities they perceive between themselves and the stimuli speakers, and therefore will likely consistently choose the most familiar accent. However, there are several other possibilities for responses to this task: they could decline to answer, say that neither speaker matches, try to conduct phonetic comparisons to find the best match for each trial, or systematically choose the Southern speakers, as a marker of non-conformity with the local majority. In fact all of these responses were given in the course of the experiment, but there was no systematic evidence that the linguistic minority sub-groups perceived themselves as speaking differently from the stimuli speakers. Several linguistic minority sub-groups are analyzed separately, in order to test whether different sets of factors influenced their similarity judgments than for their opposite sub-group or the subject group as a whole.

For children who regularly hear multiple regional accents, or a non-dominant variety of English, the association of accent with a geographical location isn't necessarily clear from the distribution of regional accents in their input. To explain the role of experience and awareness in creating the category for the local regional accent, it has to be determined on what basis children are making their similarity judgments. When the differences in children's backgrounds

accounted for, it will provide a more nuanced understanding of how experience shapes the formation of regional accent indices.

I hypothesize that children who are more aware of what a regional accent indexes, as measured in the Awareness Task, will be more likely to identify the local speakers in each trial as speaking like them, as they will understand that they share a hometown with the local speakers. I also expect that having more exposure to different regional varieties will increase children's awareness of what regional accents represent, and therefore also positively influence them to identify with the local speakers. However, I only expect this to be true of Caucasian and Insider sub-groups, and perhaps only Caucasian Insider subjects. Sub-groups of children with input from multiple varieties may be less willing to identify with the local speakers, and less aware of the location-accent connection, based on the conflicting input they may receive about the distribution of regional accents.

5.3 Methods

5.3.1 Participants

The same group of participants participated in this task as in the first task. Sixty-six children (35 female) aged 61 to 77 months, average age 70.4 months, or 5;10 participated in this task. The subjects had all been enrolled in Kindergarten in public school near Philadelphia, Pennsylvania, U.S.A. for five months when the study was conducted. All but nine were born in the town where the experiment took place, although of those, all but one had lived over half their lives in the town.

The parents of the subjects came from a wide variety of backgrounds. The majority of the subjects completing this task (40/60) were Caucasian. This is the ethnic group associated with the local regional accent used in this study. There were 13 African-American subjects. All of the African-American students were observed to be speakers of African-American English (AAE) by the experimenter.

Eleven bilingual subjects also completed this task. The bilingual subjects all spoke English, but had different histories of language acquisition, so that them as simultaneous or sequential bilinguals is difficult. I therefore have simply identified them as bilingual and did not attempt any further sub-categorization of this group.

Twenty-one of 60 subjects had at least one parent born outside of the region where the study took place. Of those, only seven were monolinguals. I refer to this group as the Outsiders; Insiders are subjects with both parents from the town where the child resides.

More detailed data on the subjects and the town in which they reside are given in Chapter 2.

5.3.2 Materials

The stimuli from this task were taken from the same set of recordings as in the first task. For a detailed description of the stimulus words and how they were selected, see the previous chapter.

In this task, two clips of the same word were heard in succession, one spoken by a Southern speaker and one by a local speaker. In each trial, children were to indicate which of the two speakers sounded most like the child.

The stimuli for this task were created from recordings of six Caucasian male speakers (three for each regional accent), 25 to 35 years of age, and all lifelong residents of their respective hometowns. The local speakers were all from the same town as the children. The non-local speakers were all from the same town in northern Louisiana, and speakers of General Southern American English. The fillers were recorded by two Caucasian men from Scotland, aged 23 to 24, who had been living in the United States for two years at the time or recording.

There were four different orders of trials, and no child saw the same order of words in this experiment as they did in the ABX task. The order of the speakers within each trial was also counter balanced between the four orders.

The slides used looked like this:

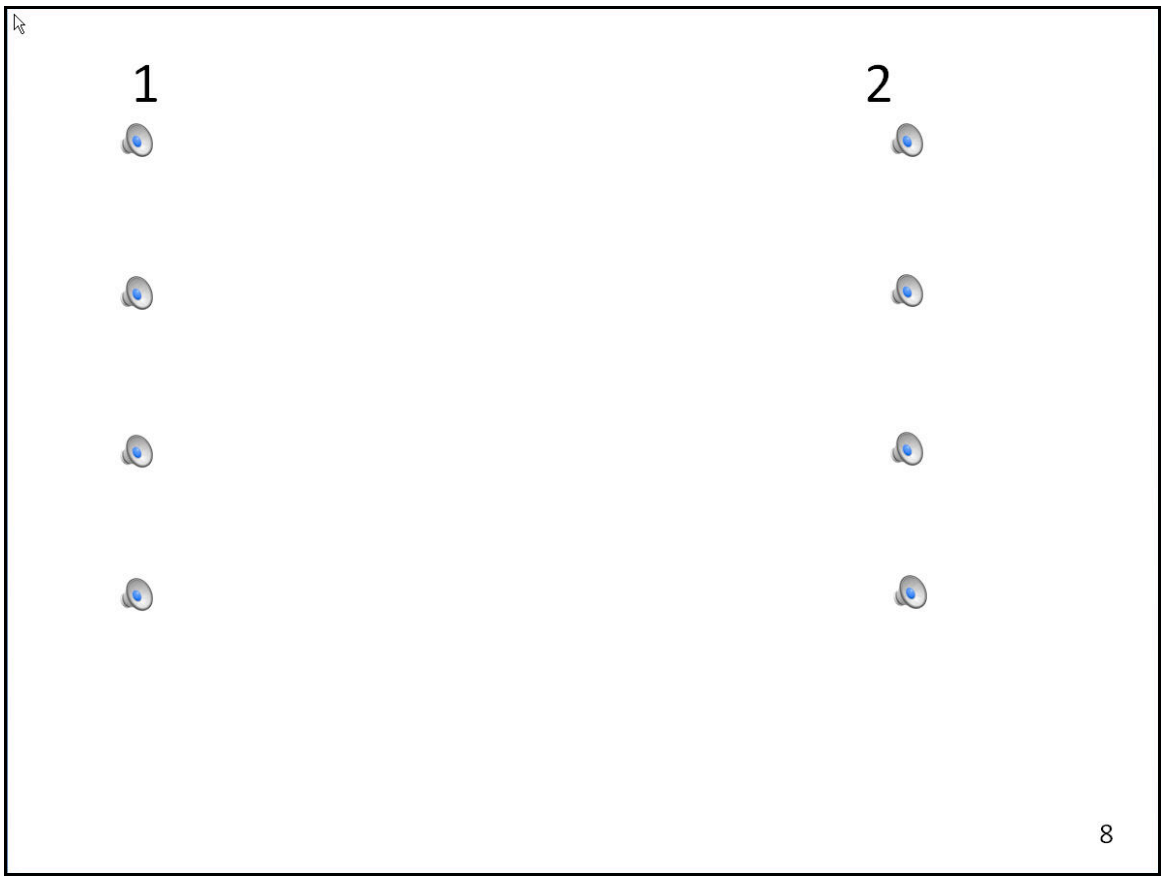

Figure 5.1 Example of Slide Used to Present Stimuli

5.3.3 Procedure

After completing the ABX task, the subjects were offered a short break, and the opportunity to stretch, use the bathroom, etc. Most of the children took the opportunity to express their interest in playing with the experimenter's computer and clicked on a few icons or pretended to type for a few minutes.

The familiarization slides with pictures of the target words were not presented a second time, since the words heard in the second experiment were the same as in the first.

The child was asked if he or she was ready to proceed, and the experimenter then explained that this time they would hear two different people say the same word, and they had to choose the one who sounded most like them. The children were told in advance that the speakers were adult males, so they wouldn't sound exactly like them, but that they should pick the best match.

The children were again given three warm-up trials, although because this was a subjective question, no feedback on whether the answer was correct was provided. The children were simply praised for completing the three trials.

Children were permitted to say that neither speaker matched their speech, to skip trials or to express reasons for their choices or reactions to the stimuli. Commentary or reaction from the children to the different speakers was recorded in a notebook, and will be discussed in the results below.

5.4 Results

In this section, I present the results of the similarity judgment task as well as of the statistical analyses looking for correlation between answers provided in the task and independent variables from the Awareness Task and Parent Questionnaire. First, the results from the entire subject group are reported, followed by each of the sub-groups. I compare each sub-group with its opposite number, so that the difference in results for the two contrasting groups can be analyzed.

The experiment consisted on 30 trials, five of which were fillers. The analyses were conducted only on non-filler trials, meaning 25 trials were analyzed per subject. I report all of the results as the number of times the local speaker was chosen. Because responses were a subjective choice on the part of the subjects, I do not refer to number correct in this task.

5.4.1 Entire Subject Group

Sixty-one children completed the similarity judgment task, and for all subjects, the average number of trials in which the *local* speaker was judged most similar to the child was 17.58 out of 25 trials, or 70% of the time. The fewest number of times a subject selected the local speaker was 7/25, the highest was in 25/25 trials. There were not enough subjects choosing the Southern speakers a majority of the time to analyze their results as a group. However, below I give a brief description of each of the subjects who systematically chose the Southern speakers.

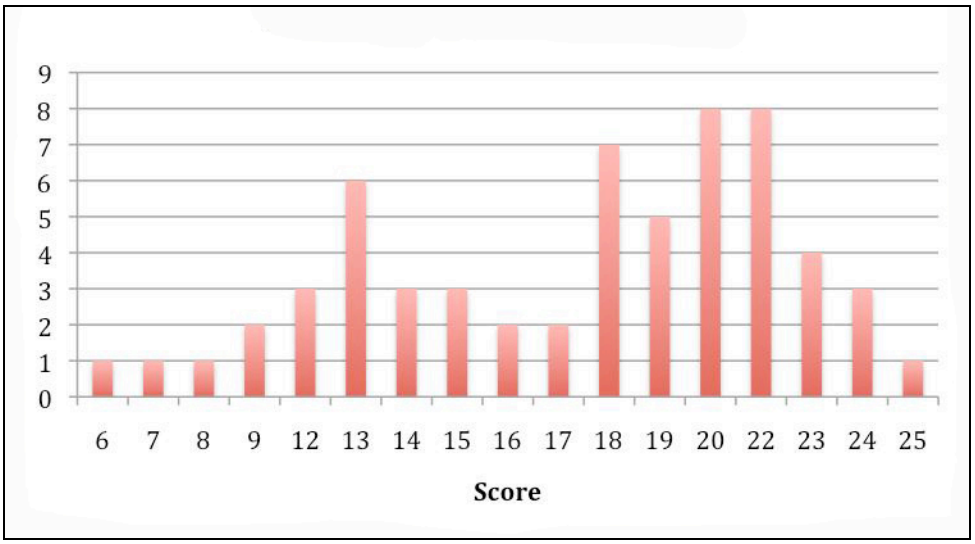

Figure 5.2 Histogram of Scores for Task 2

The most enlightening comment about how judgments were made in this task came from a subject who chose the Southern speakers in 20/25 trials. She explained to the experimenter that she chose the Southern speakers because they "talked funny" and because she went to speech therapy to learn to pronounce /r/ correctly that she also "talked funny" and therefore sounded most like the Southern speakers. Since she attended speech therapy and in that respect deviated from the rest of the subject group, her data were not included in the analyses. Her comment and results suggest that she recognized which speakers were local. However, she perceived her own speech as different enough from the majority to identify with a regional accent that she called "funny." If she is at all representative of the rest of the subject group, it is possible that five to seven year old children have very strict normative notions about what kind constitutes "normal" speech in their community. Her responses are also indicative of how other children may approach this task. Children who perceive themselves as different from the linguistic majority may express that perceived difference by claiming the Southerner is the most similar sounding speaker. Categorically judging the non-local speakers as the most similar would indicate a nuanced understanding of what a non-local accent represents: non-membership in the mainstream community. This is clearly not available to all subjects. Based on the results of the ABX task, many subjects seem to not have a clear understanding of the non-local accent and what it might represent. However, at least some children are sensitive to the fact that

language marks social affiliations, and are astute enough to recognize that in some way they are not part of the mainstream in their environment.

Four other children showed a similar pattern in their responses, and I try to attribute reasons to their choices by looking at their demographic data and their performance on the other two tasks. Two subjects chose the Southern speaker 18/25 and 17/25 times. Both of these children had scored highly on the ABX discrimination task, indicating that they had no trouble hearing the difference between the regional accents. One of the two subjects correctly identified the local accent as being local in Awareness Task Question 3, and the other correctly answered Question 5 of the Awareness Task ("why do these speakers sound different?"). When I examined the Parent Questionnaires for these two subjects, their family backgrounds provide plausible reasons for why they may have made the choices they did. One subject was a bilingual, who had one parent from Mexico, although his other parent was a Northville native. The different languages and perhaps accents in his input may have caused him to believe that he spoke differently than the local speakers, leading him to consistently identify the non-local speakers with his own speech.

The other child, who was Caucasian, had two parents from New Jersey, had himself been born in New Jersey, and regularly traveled there to visit family. Although I did not make any notes about these subjects speaking with non-local accents during the experiment, (I noted when a child spoke with any sort of accent other than the local accent represented in this task) perhaps they were aware of the other varieties spoken by their families and therefore did not choose to affiliate themselves with the local speakers. In the section on Insiders and Outsiders sub-groups below, I discuss the role of family backgrounds in greater detail.

The results of the two other children who systematically chose the Southern speakers in this task (16/25 times) were harder to interpret. The first was a bilingual, both of whose parents were from Mexico and spoke Spanish at home. This subject however, did not score well on either the Awareness Task or the ABX matching task, so it is possible he wasn't attuned differences in the regional accents in English. The other student is an intriguing case, as this child had two locally-born parents, and answered four of five of the Awareness Task questions correctly, only missing the question asking the children to identify another place besides his hometown on a map (Q2). The child did not speak another language, or belong to an ethnic

minority. He did not comment or provide me with any reasons for his choices, so I cannot surmise his motivations.

5.4.1 Statistical Analyses of Entire Subject Group

The same statistical model and procedure for selecting independent variables are used as in the ABX task. Candidate variables for analysis as fixed effects are selected via visual examination of the distribution of responses. Additionally, the two index scores, for Exposure and Awareness, are analyzed, as one hypothesis for the study is that children's ability to discriminate between regional accents improves with increased awareness of and exposure to regional accents. For more information how the index scores were created, see Chapter 3 for a description of the Awareness Task and Chapter 2 for a description of the Parent Questionnaire and the subject background data.

When the distribution of answers to items on the Awareness Task and Parent Questionnaire were examined, the following factors were determined to have the most influence on a subject's choice on any given trial: Q3, Q4 and Q5 from the Awareness Task, and the child's propensity for commenting on and imitating accents, as reported on the Parent Questionnaire. These variables are tested for their relative influence on responses in the linear regression Mixed Effects Model, the results of which are given in the table below.

Table 5.1 Correlation with Awareness and Exposure Scores

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	0.22636	0.07595	2.980	0.00288 **
Exposure	0.08708	0.09017	0.966	0.33419

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.2 Correlation with Individual Items from Awareness Task and Parent Questionnaire

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.6468	0.2693	2.402	0.0163 *
Q4	-0.5139	0.3003	-1.712	0.0870 .
Q5	0.7473	0.2915	2.564	0.0104 *
Imitate	0.5121	0.2583	1.983	0.0474 *
Comment	-0.2467	0.2611	-0.945	0.3446

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As hypothesized, awareness of regional accents has a significant, positive influence on the choice of the local speaker. Exposure, which includes interaction with speakers with accents, and watching TV programs depicting regional accents, has no significant effect on the similarity judgments.

On closer examination of the questions that comprise the Awareness Score, I found that Q3, which asked the students to say where the local speakers were from, and Q5, which asks them to more generally state why the two sets of speakers sound different, positively influence choosing the local speaker as most similar. Interestingly, the influence of having answered Q4 correctly, that is identifying the Southerners as being non-local, is negative. This means children who correctly identified the Southern speaker as not from their hometown were less likely to choose the local speaker as sounding similar to themselves. This could indicate that children who are aware that there are other regional accents in other places are also the ones who may themselves speak in another regional accent or come from families where other regional accents are used at home, and that they judge themselves as not sounding similar to the local speakers. This explanation seems to be borne out by the sub-group analyses, in particular, the Outsiders. For the entire subject group, it only is marginally significant. I suspect that this factor trends to significance for the entire subject group because of the Outsiders.

The correlation with Q3 suggests that children are aware that a particular regional accent is specific to their home region, and that they use this information to make their selection of a speaker sounding similar to them. Further supporting this idea is the correlation with Q5, in which they make a general statement about how accents are associated with geography. If

children are aware that it is a speaker's home region that determines regional accent, it is possible they are basing their similarity judgments on the knowledge that they are from the same place. This means that most subjects are not using phonetics to judge similarity between themselves and the stimuli speakers. Instead they are using socio-indexical information to judge similar speakers. This is especially interesting because the question in the task was framed "which one sounds most like me?" and not "which speaker is most similar to me?" In other words, children were prompted to compare similarity of speech and not the individual speaker. Nonetheless, they appear to be choosing similarity though based on the social qualities of the speaker and not the speech itself.

An argument could be made that a forced choice task is over-simplified, and children would naturally pick the more familiar of the two accents, regardless of how similar they actually found them. However, children had the option of saying that neither sounded similar to the child, and in fact, five subjects did respond with "neither" in some of the trials. The total number of trials in which children responded that neither speaker sounded similar to their speech was 19, which is not large enough for any further analyses. But it does show that when at least some children didn't find a similar-sounding token, they were willing to say so. I believe these cases strengthen the argument that children were making judgments of similarity based on what they know regional accent to represent, and in most cases that was the knowledge that the local speakers were members of their community. Even for the sub-groups for whom the link between regional accent and hometown was unclear, they attempted to make principled choices of similarity based on their interpretations of what regional accent indicated about a speaker, as will be shown in analysis of sub-group results below.

Two other items from the Parent Questionnaire were used as independent variables in the model: imitation of accents and commenting on accents, as reported by the parents. These two questions had been included on the Questionnaire to capture awareness of different accents children may have, even if answering direct questions about them proved difficult. It also was meant to see if children would demonstrate awareness of different accents when not being prompted by an experimenter. For the entire subject group, reported imitation of accents had a significant, positive correlation with choosing the local speaker. This correlation lends further support to the hypothesis that awareness of different accents facilitates discrimination. Commenting on accents did not reach significance in the results of the model.

To summarize the results for the entire subject group, in 70% of trials, subjects chose the local speakers as sounding similar to themselves. This again shows that children are able to discriminate between a familiar and an unfamiliar regional accent. It also shows that regional accent is a meaningful category of sociolinguistic variation for children at this age, because they are able to overcome mismatches in age and gender to use regional accent as a measure of similarity of speech. Finally, it also demonstrates the children understand and can access the socio-indices of the local accent, and use this information to identify similar sounding speakers.

For children speaking the local majority regional accent, this is perhaps an easy task, as they in theory could test for similarity by comparing the stimuli to their own productions. However, for subjects from different ethnic and linguistic backgrounds, their history of input in other varieties and affiliation with other speech communities may make the choice difficult. The fact that subjects belonging to linguistic minorities consistently chose local speakers as sounding most similar suggests that at least some subjects are making similarity judgments based on what they interpret regional accent as representing (shared community, or shared area of residence), rather than any phonetic similarity in their speech.

The set of independent variables that positively influences selection of the local speakers as similar is related to how aware the children were that regional accents are linked to where a speaker lives, and in particular whether that speaker is from their hometown. I submit that, based on the significant influence of knowledge about the meaning of regional accents, and that a particular regional accent marks membership in the local community, most of the subjects in this study used socio-indexical information to make their similarity judgments rather than phonetic comparisons.

Finally, the lack of influence that any exposure variables have on responses of the entire subject group was unexpected. However, the reported amount and intensity of exposure to regional accents were subject to the parents' interpretation and accuracy of their reporting. They also were not asked to follow children's exposure, but answer a single set of questions based on their memory of past exposure. Therefore, more precise measures of exposure or more in-depth case studies may be needed to fully assess the effects of these factors on ability to discriminate regional accents.

5.4.2 Sub-groups

Sub-groups of subjects were analyzed to see whether the pattern of correlating factors varied by subject's linguistic background. Although most of these sub-groups judge the local accent similar to their own speech with the same frequency as the group as a whole, differences in exposure to other varieties and the experience of being a minority-variety speaker may differentially shape a child's interpretation of regional accents.

The sub-groups were created in pairs from the entire group of subjects: Monolinguals and Bilinguals, Insiders and Outsiders (children with 1+ parents from Northville, and children with both parents from outside of Northville, respectively) and African-Americans and Caucasians. Other ethnic groups were represented in the study, but not in great enough numbers to conduct analyses on their performance as a group. These groups were chosen based on visual inspection of the distribution of their similarity judgments plotted against independent variables such as responses to Awareness Task questions or demographic variables.

Below I present the results from each pair of sub-groups, followed by a discussion of the findings.

5.4.3 Insiders and Outsiders

For this analysis, the subjects were broken into two groups, based on whether at least one of the children's parents was born in Northville (Insiders) or whether both parents were from outside the region (Outsiders). Children of a single parent were grouped based on the place of birth of that parent. Only monolinguals were included in these two sub-groups, in order to try and isolate the effect of hearing other regional varieties (as opposed to non-native ones) at home.

The logic behind this grouping is that a child with both parents from outside the region is likely to have intense exposure to a non-local variety, in addition to the local variety they hear in the community. They may have significant experience traveling to visit friends and family in that other region, and may interact on a regular basis with people from that region, and perhaps hear more commentary on the differences in speech between their home region and that of the parent, or feel affiliation with that other region, based on the family's association with it.

This of course does not preclude a child with locally-born parents from having similar experiences, but the daily exposure to another regional variety from a parent will still be missing. The intense exposure to another regional variety at home affects how children, born in that

particular area, produce sounds specific to the local variety, as shown by Payne (1976). That study, which was conducted approximately five miles from where the present study took place, albeit 30 years earlier, showed that Outsider children often did not acquire some of the more subtle regional phonetic markers in their speech. Only children of parents indigenous to the region fully acquired the local variety. Although that study dealt with production, and this one subjective perception, it seems reasonable to expect that having non-local parents might affect one's perception of regional accents.

An additional piece of evidence supporting the idea that being an Outsider might affect ability to hear regional accents is found in Floccia, Girard et al (2009). In this study, seven year olds with parents from outside the region where the study was conducted were more accurate in categorizing regional accents than children with autochthonous parents.

The average number of trials in which Outsiders (n=10) picked the local speaker was 16.6/25 (66%). The average number of times the Insiders (n=41) picked the local speaker was 18/25 (72%). This was not a significant difference*: $t=0.2585, p=0.797$* in a two-tailed t-test and *$W=448, p=1.00$* in a Wilcoxon Rank Sum Test. The averages of the two sub-groups were not significantly different whole subject group either. (Insiders *$W=840, p=1.00$* in a Wilcoxon Rank Sum Test*, $t= 0.163, p=0.8711$* in a two-tailed t-test; Outsiders *$W= 960, p=1.00$* in a Wilcoxon Rank Sum Test, *$t=-0.1363, p=0.892$* in a two-tailed t-test). As in the previous task, the results between the two-tailed t-test and the Wilcoxon Rank Sum test converge.

The factors influencing those choices differ drastically between the two groups, likely the result of exposure to regional variation in the home. The fixed effects correlating with similarity judgments are examined for each sub-group separately and then compared and discussed below.

5.4.3.1 Insiders

The Insiders show the pattern seen both in the entire subject group as well as in several other sub-groups: the overall Awareness Score, as well as Q3, identifying the local speakers as being from Northville, and imitation of accents, as reported on the Parent Questionnaire, all positively influence the choice of the local speakers in this task. Q5 of the Awareness Task (stating the general reason why the two sets of speakers sound different from one another), is marginally significant for the Insiders, but is still in the positive direction, as with the subject group as a whole.

The Insiders sub-group was the only one that shows an effect of Exposure on their responses, and as it turns out, travel is a significant component of that measure. Both the Exposure Score and having traveled to other regions have positive correlations with judging the local speaker similar to the child's own speech.

Considering the background of the Insiders, it could be that travel to other regions is what makes those children aware that regional variation exists. I suspect that children with parents from the same hometown get a disproportionately large amount of exposure to the local regional accent as compared to Outsider children. Therefore, for Insiders, leaving the hometown may be the catalyst for realizing regional accent is a kind of social variation. The experience of traveling to a different place and hearing a different variety makes the connection between regional accent and location clear.

Table 5.3 Correlation with Awareness and Exposure Scores for Insiders

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	0.26813	0.06838	3.921	8.81e-05 ***
Exposure	0.25507	0.10308	2.475	0.0133 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.4 Correlations with Individual Items for Insiders

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.6347	0.2619	2.424	0.01537 *
Q4	-0.1049	0.2900	-0.362	0.71746
Q5	0.5041	0.2845	1.772	0.07638 .
Imitate	0.8237	0.2698	3.053	0.00226 **
Comment	-0.4984	0.2716	-1.835	0.06648 .
Travel	0.6940	0.2514	2.760	0.00578 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.4.3.2 Outsiders

In the Outsider sub-group, all the subjects had the same response for the Comment and Travel variables. Because there was no between-subject variation in these two factors, it was impossible to use these two factors together as fixed effects in the same analysis. I therefore ran the analysis twice, using these two variables in turn as fixed effects. The results, of course, are identical regardless of whether Comment or Travel is used as a fixed effect in the model. However, in order to compare the effects to the Insiders, I wanted to have both variables accounted for in the analyses.

The Outsiders show almost an opposite pattern of fixed effects reaching statistical significance from the Insiders. In this case, recognizing the Southerners as non-local (Q4) and commenting on other accents have a significant negative influence on the choice of the local speakers as similar. That is to say, the better they are at realizing the Southerners were non-local, and the more frequently they comment on accents and travel, the less likely they are to choose the local speakers as sounding similar to themselves.

The only significant, positive influence on their choice of the local accent as similar is knowing that the two sets of speakers differ mainly because they are from different places (Q5). All other significant factors have a negative correlation.

I suspect that these correlations reflect how Outsider children make their similarity judgments, and how they are bringing their experiences with different regional accents to bear on this task.

Outsiders are children with both parents originating from outside of Northville. They have received conflicting evidence about which accents are local, in that they hear both the majority regional accent and another regional accent in their environment. This effectively makes both regional accents local to them, unless they understand their parents are not from the same town originally as they. The child may hear family members with regional accents similar to the parents, and they may feel affiliation with these speakers. Thus, for Outsider children, the distribution of local people speaking one variety versus non-local people speaking a different variety isn't as clear as it is for most Insider children. Outsiders' experiences with different regional accents complicate the recognition of local speakers via socio-indexical information provided by the accent that other sub-groups use to make similarity judgments.

Having commented on accents in the past also had significant negative influence on the choice of the local speaker. This may point to the possibility that those children don't have a firm idea about how regional accents are distributed, and are testing theories about it out loud, perhaps seeking feedback. Most parents did not provide specifics on the kinds of comments the children made about accents, but it could be commentary is an expression of confusion or an attempt to figure the regional accent puzzle out, and not a measure of awareness, as I had initially expected. More work will be needed before the role of children's commentary can be accurately determined.

Travel also showed a significant negative correlation with choosing the local speaker for the Outsider sub-group. Increased travel may represent greater exposure to the non-local accent, and perhaps also affect an Outsider child's sense of affiliation with the non-local accent. Again, ethnographic work on subjects would help to explain this effect, which was in the opposite direction of what I initially hypothesized for this factor.

Of the 10 subjects in the Monolingual Outsiders sub-group, seven answered Q4 incorrectly. For Q5, 6 of 10 subjects (but not the exact same group who answered Q4 incorrectly) answered Q5 incorrectly. Two subjects answered both correctly and five subjects had both Q4 and Q5 wrong. Of the remaining three subjects, one had Q4 correct and Q5 wrong and two had the opposite pattern of results.

I have broken the results down by subject to show that a small sub-group of monolingual Outsiders may have been responsible for the conflicting direction of correlations found in this analysis. One small group, who knew the Southern speakers were non-local and identified the Southerners as sounding similar to themselves, may have also known that their families were also non-local. As a result, they may believe that they do not speak like the majority of people in their community, or are not fully socially integrated into this community. All of the subjects who answered Q4 correctly were Caucasian, meaning race likely did not play a role in their similarity choices.

One of the subjects who answered Q4 correctly almost categorically chose the Southern speakers as similar, (perhaps driving most of the effect found here). She also correctly answered Q3, so the choices made in this similarity judgment task were not the result of erroneous identification of the regional accents. A more likely explanation for her pattern of similarity judgments is that she was marking a perceived difference between herself and the rest of the

community. This is not an unlikely explanation; another subject, mentioned in Chapter 3 and section 5.4.1, and who was later excluded from the analyses, categorically chose the Southerners as similar sounding and explained to the experimenter this was because she had a speech impediment and also spoke "funny," like the Southerners.

The reasoning behind subjects' choices in this task cannot be definitively explained from the results of this study, but with more in-depth questioning of the children and parents, might be possible to establish in future studies. The results of the Outsiders underscore the importance of collecting extensive background data on subjects in studies dealing with accent perception, and suggest a need to debrief subjects about their choices on similarity judgment tasks.

The take-home point is that children may use accent to mark social difference in this task between themselves, their families and the community by claiming similarity with another regional accent. This in turn suggests that they have a rather sophisticated understanding of how an accent marks in-group and out-group membership, and where they stand vis-à-vis that boundary within their community. Phonetically speaking, their judgments are not correct (none of the children had a noticeable regional accent, based on my interactions with them), however, are quite sophisticated in their understanding of socio-indexicality.

The positive correlation between Q5 and identifying with the local speaker for the Outsider sub-group lends itself to a more straightforward explanation: Outsider children who know how regional accents are distributed have an easier time choosing the most similar sounding speaker. This was also true (although the correlation was only marginally significant) for the Insiders. Outsiders and Insiders answered Q5 at approximately the same rates (38% vs. 39%, respectively) but the Outsiders drew on this information much more heavily when making their similarity choices. It could be that those subjects were using that information to explain any differences they perceived between their own (and by extension, family's) speech and the majority speech variety in the community.

The correlation found with Q5 also provides us information about Outsiders unaware of regional accent distribution. Children who did not answer Q5 correctly were less likely to choose the local speakers as sounding similar in this task. I suspect this is because the connection between location and regional accent is unclear for some Outsiders, a result of their linguistic experience at home. The knowledge represented by Q5 may play a more important role for Outsiders, as compared to other subjects. They were the only sub-group in the ABX

Task that did not show a significant negative effect of Q5 on discrimination, perhaps indicating that rather than complicating discrimination, this information helps them to understand the distribution of accents in their environment.

Table 5.5 Correlations for Awareness and Exposure Scores, Outsiders

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	-0.2415	0.3217	-0.751	0.4529
Exposure	-0.3591	0.3053	-1.176	0.2394

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.6 Correlations with Individual Items, Outsiders

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.49931	0.34813	1.434	0.15149
Q4	-1.95048	0.41097	-4.746	2.07e-06 ***
Q5	0.93704	0.42152	2.223	0.02622 *
Imitate	0.05317	0.39285	0.135	0.89234
Comment/Travel	-3.26204	1.12331	-2.904	0.00368 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The Insider and Outsider sub-groups are the most interesting of all those compared. Their backgrounds affect the sets of strategies used to complete the task, but nonetheless they have similar rates of choosing the local speakers in this task. For the Insider children, the distribution of regional accents is clear, and they use this knowledge to identify similar sounding speakers using socio-indexical information associated with the local accent. For many Outsiders, the connection between regional accent and shared community is not as obvious. However, one factor seems to help them reason through the distribution of accents in their environment: the knowledge that regional accents have to do with geographical distribution makes it more likely that they affiliate themselves with the local regional accent.

Finally, several Outsider children have shown a grasp of the socio-indexical nuances of regional accent in this task. Children belonging to the racial majority in the town, but who perceived some difference between themselves and the local population, chose the Southern accent as sounding most similar. I interpret this as a sophisticated understanding of how speech marks group membership. They have grasped the iconicity of an accent, and what it says about the speakers. Although those children are from the town and know exactly what the local accent sounds like, there is something that causes them to not want to affiliate themselves with it. They may not know what the indices are of the accent they did judge as similar to themselves, but it is clear they have a detailed understanding of what the local accent represents in the context of their hometown.

5.4.4 Caucasians and African-Americans

5.4.4.1 Caucasians

The Caucasian sub-group (n=36) chose the local speakers in 18.3/25 (72%) trials, not a significant difference with either the African-American subjects (n=13): two-tailed t-test: *t=-0.9473, p=0.3534*; Wilcoxon Rank Sum *W=200, p=0.217*, or the subject group as a whole *t=0.7921, p=0.4305* in a two-tailed t-test; Wilcoxon Rank Sum *W=1083, p=0.41*.

Table 5.7 Correlations with Awareness and Exposure Scores, Caucasians

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	0.09948	0.11998	0.829	0.4070
Exposure	-0.02824	0.12065	-0.234	0.8149

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.8 Correlations with Individual Items, Caucasians

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.4035	0.4254	0.948	0.34289
Q4	-0.6531	0.3804	-1.717	0.08596 .
Q5	0.7910	0.3348	2.363	0.01814 *
Imitate	0.5801	0.3181	1.824	0.06818 .
Comment	-0.8898	0.3282	-2.711	0.00671 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.4.4.2 African-Americans

The African-American sub-group (n=13) chose the local speaker as sounding similar to their own speech in 17.1/25 (68%) trials on average (the whole subject group did so for 17.6/25 trials, which was not a significant difference in a two-tailed t-test, *t=-0.4018, p=0.6921* or a Wilcoxon Rank Sum Test, *W= 437.5, p=0.496*). All of the African-American subjects, however, were speakers of African-American English (AAE), as judged by the experimenter during initial conversations with each subject. There was therefore a real possibility for this particular sub-group to not identify with the local majority regional accent, usually associated with Caucasians in this context.

Table 5.9 Correlations with Awareness and Exposure Scores, African-Americans

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	0.30267	0.08399	3.604	0.000314 ***
Exposure	0.12465	0.12036	1.036	0.300365

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.10 Correlations with Individual Items, African-Americans

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.8949	0.2872	3.116	0.00184 **
Q4	-0.8029	0.5919	-1.357	0.17495
Q5	1.4192	0.6616	2.145	0.03195 *
Imitate	-0.2249	0.3645	-0.617	0.53723
Comment	1.1403	0.3608	3.160	0.00157 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The results of the Mixed Effects Model show different sets of correlating effects for each of the sub-groups. For the African-American sub-group, the Awareness Score has a significant, positive correlation, as does Q3, identifying the local speaker as being from Northville, and Q5, correctly stating that regional accent was the difference between the two sets of speakers. Following the argument for the subject group as whole, I suspect that these subjects were making their similarity choices based on knowledge that the regional accent of the local speakers marked them as members of their community, and not phonetic similarity. If the children recognized a speaker as being from the same town as they, they were willing to judge that person's speech similar to their own by extension.

None of the other individual Awareness Task questions or any of the exposure items from the Parent Questionnaire correlated. However, unlike with the entire group, parental reports that the children comment on other accents has a significant, positive influence on responses in the experimental task. This may be indication of awareness of accents in interactions with others.

As for the Caucasian subjects, a slightly different pattern emerges. Neither Awareness nor Exposure Scores correlate with the children's responses. However, as with the whole group, Q3 and Q5 positively predict choosing the local speaker, as does imitating accents. Both Q4 (identifying the where the Southern speakers are from) and commenting on accents as reported in the Parent Questionnaire show significant negative correlations.

Note that the direction of the correlation for commenting on accents is different for African-American and Caucasian subjects. Caucasian subjects (as well as Outsiders) show a negative correlation between commenting and choosing the local speakers, whereas the African-

Americans have a significant, positive correlation. I have no additional data to explain the difference between these sub-groups, but investigating the kinds of commentary children make about regional linguistic variation is a priority for future work.

The most surprising result of this sub-group comparison is that the African- American and Caucasian subjects have the same frequency of selecting the local speakers as similar. I hypothesized that African-American subjects might have several options in making similarity judgments in this task. I thought it likely that they would either not judge either speaker in this task to be similar sounding, or that they would systematically choose the Southern speakers based on perceived dissimilarity with the local speakers, or actual phonetic similarities with Southern American English.

The pattern of correlations found here indicates that children use the socio-indexical information provided by regional accents. The African-American sub-group demonstrates that children can perceive regional variation and socio-indices in dialects that are not their own. They also seem to use the information provided by the accent about the speaker as a means of finding similarity, as opposed to comparing their speech to the stimuli, or at least give this information preference when making similarity judgments. This is especially interesting given the salience of ethnicity, both in terms of appearance and its association with linguistic variation in this context and in United States culture in general.

The results from these two sub-groups suggest several possible avenues for future research. Conducting a matched-guise experiment with children in this age group would provide insight into whether they, like adults, report hearing accents when cued by non-linguistic information, given their lack of attention to ethnicity when making similarity judgments in this task. It would be beneficial to fully understand which varieties children designate as local, and how experience with different ethnolects in the community shapes this tendency. Further experiments should test whether Caucasian children are willing to judge speakers of a local AAE variety as being similar based on speech, or whether both African-Americans and Caucasians would recognize Hispanic English, another variety found in that community, as similar to their own. These experiments are planned for future work.

5.4.5 Monolinguals and Bilinguals

These two sub-groups were the only two that showed significant differences in their rates of selecting the local speaker as sounding similar. The Monolinguals were, in fact, the only sub-group that picked the local speakers at a different rate than any sub-group or the entire subject group; they were significantly more likely to choose the local speaker as sounding similar to themselves.

The Bilinguals, n=11, averaged 15.9/25 trials selecting the local speakers, approximately 64% of the trials. The Monolinguals, n=49, averaged 21.3/25 (85%) trials in which they selected the local speaker. This was a significant difference in a two-tailed t-test (*t= -3.1146, p= 0.007436*) and in a Wilcoxon Rank Sum Test (*W=114.5, p=0.0031*).

As compared with the entire subject group, the Bilinguals were not significantly different (*t= -0.9862, p=0.3419* in a two-tailed t-test, and *W=387, p=0.3674* in a Wilcoxon Rank Sum Test). Monolinguals were significantly more likely than the subject group as a whole to pick the local speaker (*t=4.0073, p=0.000119* in a two-tailed t-test; $W=830, p=9.47 \times 10^{-5}$ in a Wilcoxon Rank Sum Test.)

The Monolinguals pattern with the entire subject group in their correlations was unsurprising given that they comprise such a large portion of that group, 49/60 subjects. They show a significant, positive effect of answering Q5 (a general knowledge of why the two speakers with regional accents sound different) correctly, and a marginally significant positive effect of Q3 (correctly identifying the local speakers as being from Northville). Similarly, the Awareness Score positively correlates with selecting the local speakers as similar as does imitation reported on the Parent Questionnaire. This was the only sub-group for which imitation of accents reached significance.

The Bilinguals, however, only have one factor that reaches significance in the analysis, and that is Q3. It has a positive influence, meaning that these subjects are also likely using knowledge that the local regional accent indicated membership in the community, to make their choices. No other factor reaches significance.

Note that none of the bilinguals correctly answered Q5 in the Awareness Task, so it could not be included in the analysis, since variables must have binary values in order for the model to calculate their levels of relative influence.

5.4.6.1 Bilinguals

Table 5.11 Correlations with Awareness and Exposure Scores, Bilinguals

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	0.1170	0.2331	0.502	0.616
Exposure	0.3753	0.2327	1.613	0.107

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.12 Correlations with Individual Items, Bilinguals

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	1.3078	0.6043	2.164	0.0304 *
Q4	0.1148	0.7436	0.154	0.8773
Q5	N/A			
Imitate	-1.1359	1.1382	-0.998	0.3183
Comment	1.7815	1.1538	1.544	0.1226

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.4.6.2 Monolinguals

Table 5.13 Correlations with Awareness and Exposure Scores, Monolinguals

	Estimate	Std. Error	Z Value	Pr(>|z|)
Awareness	0.21331	0.08352	2.554	0.0106 *
Exposure	0.03467	0.09693	0.358	0.7206

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5.14 Correlation with Individual Items, Monolinguals

	Estimate	Std. Error	Z Value	Pr(>|z|)
Q3	0.5442	0.2926	1.860	0.06286 .
Q4	-0.4536	0.3158	-1.436	0.15097
Q5	0.8207	0.2952	2.780	0.00543 **
Imitate	0.6463	0.2670	2.421	0.01549 *
Comment	-0.5157	0.2747	-1.877	0.06053 .

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.5 Item Analysis

The linear regression Mixed Effects Model was also run using the item (the word heard in each trial) as the independent variable and subject as a random effect (in the previous analyses, both the item and the subjects were treated as random effects.) This was done in order to see if any items were significantly more difficult or easier than others.

The results, shown below, indicated that two of the items were significantly easier than the rest, "buy" and "pie." These belong to the PRICE (following Wells (1982)) class of words, which were also the easiest for subjects to match in the ABX task, so it is expected that these items are also salient in this task.

Table 5.15 Effect of Item

Word	Estimate	Std. Error	Z Value	Pr(>|z|)
Boat (Intercept)	0.75928	0.30865	2.460	0.01389 *
Buy	0.96151	0.45330	2.121	0.03391 *
Coat	0.26223	0.41263	0.636	0.52510
Day	0.31004	0.41790	0.742	0.45814
Goat	-0.24886	0.40347	-0.617	0.53736
Hay	-0.16599	0.40085	-0.414	0.67880
Heel	0.35440	0.41617	0.852	0.39444
Juice	0.35440	0.41617	0.852	0.39444
Light	0.61086	0.43180	1.415	0.15716
Neigh	0.12464	0.41471	0.301	0.76377
News	0.84332	0.44475	1.896	0.05794 .
Night	0.75854	0.43656	1.738	0.08229 .
Note	-0.24886	0.40347	-0.617	0.53736
Peel	0.36199	0.42152	0.859	0.39047
Pie	1.73425	0.53360	3.250	0.00115 **
Play	0.62109	0.43126	1.440	0.14982
Sail	0.22885	0.41374	0.553	0.58017
Seal	0.65103	0.43030	1.513	0.13029
Sell	-0.60894	0.39762	-1.531	0.12566
Tail	-0.08374	0.40256	-0.208	0.83522
Tie	0.84332	0.44475	1.896	0.05794 .
Tune	-0.36974	0.39962	-0.925	0.35485
Well	-0.03890	0.40573	-0.096	0.92362
Wheel	-0.24704	0.39942	-0.618	0.53625
You	0.32149	0.41722	0.771	0.44097

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.6 Discussion

This task asked children aged five to seven from a variety of different linguistic and ethnic backgrounds, but all residing in the same community, to choose the speaker sounding most like themselves. This was done to see whether children can use the social indices of regional accent to identify similar sounding speakers, even when faced with differences in age, gender and ethnicity between themselves and the speakers. I also tested whether increased awareness of regional accents or exposure to different kinds of accents affects the likelihood that children will identify other speakers from the same home region as sounding similar to themselves.

The subjects all systematically picked out one regional accent from the set of speakers, indicating that they could all hear the differences in regional accent. Seventy percent of the time, the subjects chose the majority local regional accent in response to the question "which speaker talks most like you?"

This in itself was remarkable, as I expected subjects from minority linguistic backgrounds to either not systematically pick the local speakers, or to respond that neither speaker sounded like them. All speakers used in recording of the local accent stimuli were Caucasian adult males, speaking the most prevalent (Caucasian) variety spoken in that town. Therefore, the tendency for the majority of subjects to affiliate themselves with the local speaker indicates that they were aware of the social indices of regional accent, that it marks a speaker's hometown, and that they used this information to choose a similar speaker. Put another way, subjects ignored actual differences in their speech from what was spoken in the stimuli and based their similarity judgments on the fact that they were from the same place as the speaker.

The question immediately arises how children were making their judgments. Were they making comparisons to their own speech, were they simply choosing the most familiar accent, or were they utilizing their knowledge of the social index of the regional accent? However, because extensive background information was collected about the subjects and about their knowledge of regional accents, I was able to test for the influence of this knowledge on their responses in this task, and determine if it influenced their responses. Based on correlations found in a Mixed Effect Model, most subjects appear to have been positively influenced to choose the local speakers as most similar when they knew that regional accent was from their hometown. That is

to say, they may have identified the local speakers as similar to themselves based on the knowledge that those speakers were from the same town as the child.

With the exception of the subjects from Outsider families, i.e., those in which the parents were not originally from the Northville area, the most frequent correlation was when children knew that regional accents marked where a speaker was from; they were likely to choose the local speakers as sounding most similar to themselves. This was true even for the Outsider subjects. This suggests that the explicit awareness of what a local regional accent represents had a powerful effect on the children's similarity judgments. The second most frequent factor was having correctly identified the local accent as being from Northville was the most common factor predicting the choice of the local speakers in this task across the sub-groups.

This finding strongly suggests that children have developed an exemplar category of local speech by this age, and the category has socio-indexical information associated with it. Regional accent appears to be a well-entrenched variety in their understanding of linguistic variation, and they are not confused when making their judgments by differences in age, gender or ethnicity between themselves and the stimuli speakers.

The results also suggest that this well-established category for "local regional accent" allows children to use non-linguistic information to make linguistic judgments, a phenomenon well-documented with adults. This is especially interesting, since studies of American adults show that they too don't seem to focus on phonetic details that distinguish sub-groups of regional accents (Clopper & Pisoni, 2004a) and that perception of phonetic markers of regional accents can be manipulated by telling listeners that the speakers are from a particular place (Niedzielski, 1999). In this experiment, I don't believe that children considered the phonetic similarities with the speakers, beyond what they needed to identify them as local or non-local, and in most cases, based their judgments on the fact that they were from the same place as the speakers. As in the first experiment, they use the social-indexical information about the speakers to make their choices.

At least half of the subjects had reason to not claim to sound like the local speakers in this task, based on their linguistic backgrounds. However, only 5 subjects systematically chose the Southern speakers in this task, and in fewer than 20 trials (out of 1500) did subjects not find either speaker to be a good match. Children belonging to linguistic minorities were willing to overlook actual phonetic differences in the face of socio-indexical information about similarity.

The most interesting finding of all deals with the role of experience in this task. Here the children show different effects of intense personal experience with regional accents. Those children (here dubbed "Outsiders") who hear a different regional accent at home, and who travel to regions with different accents, are aware that they hear multiple regional accents in their local environment. The likelihood that they choose the local speaker as similar improves when they are explicitly able to state how accents are distributed in Q5. Regional variation requires additional information in order to correctly attribute it to its source. Children who are aware of the source of regional variation use this to help identify the local speakers in this task.

In the ABX task, the correlation between Q5 and correctly matching speakers was negative. In this task, the direction of the correlation is positive. I suspect this is the case because children are not asked to interpret the non-local accent, as they are in the ABX; it suffices to recognize the local accent in order to choose the most similar sounding speaker. In this task, they are not trying to draw on general knowledge about accents to interpret the Southern accents, and the information in Q5 serves to reinforce their choice (in most cases).

It is mainly in the Outsider sub-group with children claiming similarity with the Southern speakers, although they have correctly identified both the local and non-local speakers in the Awareness Task. This may be an affective choice, but shows an extreme sensitivity to what a local accent means in their context.

The question that now remains to be asked is whether children are aware of how they themselves speak. That is, are children with minority accents (such as ethnolects or other regional varieties) residing in that community, aware that they do not speak with the majority regional accent? This was not directly tested in this study, but should be included in debriefing questions in future studies, to understand if they know about their own speech variety.

Further exploration of their representations of local regional accent also must be conducted to determine whether children are equally accepting of other kinds of accent they hear in their environment as local. For example, would Caucasian children accept AAE as local, as it is also frequently heard in their hometown?

A final question arising from this task is whether the experimenter biased the subjects' judgments in this task, and whether this should be controlled for in future experimental work. Although I do not speak with the most common Caucasian regional accent in that town, I am Caucasian and it cannot be ruled out that this fact may have biased some minority subjects. It is

not unprecedented that seeing or interacting with other nationalities biases speech perception (Hay & Drager, 2010; Hay, Warren, & Drager, 2010), and speech accommodation is well documented between interlocutors with different accents and varieties (Howard Giles, Coupland, & Coupland, 1991). Therefore running this study with an experimenter of another race may well affect the outcome for some non-Caucasian students.

5.7 Conclusion

The results of this task show that most children in the study were able to recognize one set of speakers as local, and utilized this knowledge to systematically select the local speakers as sounding most similar to themselves. This was the case even for many subjects whose families were not speakers of the majority variety in that town. The results indicate that the children were making their choices not based on phonetic similarity, but on the knowledge that they were from the same region as the local speakers. The correlations between responses in this task and the subject background and awareness data provide insight into the kinds of information children use in recognizing other speakers from their hometowns. These results suggest that children use socio-indexical information about speakers to make judgments of speaker similarity, even in the face of conflicting phonetic information. It seems that not only are they capable of interpreting the indices of accent, but that they use them when judging the similarity of other speakers' speech.

CHAPTER 6

General Discussion

Below, I summarize the findings and results of the three tasks as they relate to the theoretical framework and empirical problems discussed in the first chapter of this dissertation.

6.1 Empirical contributions

One of the motivations for this dissertation was to show that children ages five to seven can discriminate between regional accents, as previous work claimed that they had difficulty with this task (Floccia et al., 2009; Girard et al., 2008). I believe that the reason that they succeeded in these tasks but have failed in previous ones is a direct result of how the tasks were framed. Children's understanding of what regional accent is and how it is distributed is likely not fully developed; they don't have the experience with the range of accents, or exposure to the stereotypes of regional accents that an adult does. This means they may not understand the connection between regional linguistic variation and a particular location. As a result, instructing them to group speakers by accent or location may seem like an arbitrary task. Not explicitly referencing accent and its connection to location avoids this potential source of confusion, and therefore provides a clearer picture of whether children can recognize and interpret regional accents.

Secondly, framing of the task can bias how children approach discrimination, and what sociolinguistic variables they listen for. In this study, I asked how children would perform on regional accent discrimination when they weren't explicitly told to listen for regional variation. I also ask children to discriminate between speakers by referencing a social characteristic of themselves, to see whether region of residence was the preferred social characteristic with which to identify similar sounding speakers. The two tasks fundamentally ask the children to do the same thing: discriminate between two regional accents. However, in both tasks, many children

use regional accent as the basis for matching, and in some cases there is correlation between identifying the local regional accent and making correct matches.

In the ABX Task, this correlation was only found in trials where speakers with the local accent matched. The children's performance also did not correlate with any exposure to different accents or knowledge of regional linguistic variation.

In the second task, which specifically asks children to reference their own speech and make a social judgment about the speakers' regional provenance, there are more correlations between the meta-linguistic awareness of regional accents and the responses they provide in the experimental task. Note that this second task, like the ABX task, also does not explicitly reference accent or locations. But asking children to judge whether another person is similar to themselves is asking for a social judgment, and encourages the children to consider that information when responding.

As a result, there are many more correlations between Awareness Task questions and responses to the second discrimination task. It is possible that by directing the children to use the social information encoded in the accents in this task, they draw on it much more heavily than when no suggestion of socio-indexicality is made. The difference in how meta-linguistic awareness of regional accents correlates with the responses on the two discrimination tasks supports my claim that framing of the tasks matters. Neither task referenced accents or locations, but yet there were strong correlations - positive and negative - between knowledge of regional accents and performance on both discrimination tasks.

The mixed directions of the correlations in the two different tasks confirm that the development of socio-indexical knowledge is not straightforward or uniform, and is very much dependent on a child's linguistic background. When experiments on regional accent discrimination are conducted with adults, the assumption is made that the subjects have the same social categories and understanding of regional accents as the experimenter. Even with adults, perception has been shown to be heavily dependent on speaker experience hearing those accents (Clopper & Pisoni, 2004a; Sumner & Samuel, 2009). It is therefore critical when conducting experimental work with children that their exposure to regional variation is taken into account, and that no assumptions are made about how they perceive regional variation.

Finally, the experimenter cannot assume all children even have the same understanding of what the local accent is. It appears that their experience hearing different regional accents at

home and in the community affects how they define the local accent(s). This underscores the importance of collecting and analyzing data about individuals participating in the experiments, and the contributions that experience makes to forming regional accent categories.

In sum, when studying accent discrimination with children, care must be taken to assess the baseline levels of knowledge and familiarity that children have with the accents used in the study. Children understand familiar and unfamiliar accents in different ways, and the effects of their conscious knowledge about accents affects their interpretation of familiar and unfamiliar accents differently. Finally, experimenters should also be aware that referencing accents may be misleading to young subjects, since they have highly divergent understandings of what an accent is, dependent on their experiences with different kinds of variation.

6.2 Theoretical Contributions

Exemplar Theory has increasingly been used to explain the effects of social knowledge on speech perception, usually positing the existence of abstracted categories, built from memories of perceiving speech, in which phonetic/acoustic information about speech is linked to information about the speaker's social identity. Below I address several predictions made by other researchers of exemplar theory about how exemplar category formation might take place in children. I compare the results of this dissertation to those predictions to see how well they fit or what changes might be made to the model in order to accommodate my findings.

6.2.1 The Formation of Exemplar Categories of Social Variation

Foulkes and Docherty (2006) make two predictions about how the creation of exemplar categories for social variation would proceed: those social categories that are most frequently encountered and that are most transparent (e.g. gender, ethnicity) would be the first to be acquired. Social categories that are not transparent, or are arbitrarily defined, would be last to be acquired. Regional accent would fall, based on their predictions, on the more difficult end of the scale of acquisition.

These two predictions are borne out by the findings of this study. First, the most frequently encountered category, local regional accent, appears to have a more robust representation than the less-frequently encountered category, the non-local regional accent. Regional accent is not a well-entrenched category for all of the subjects in this experiment.

There was quite a bit of variation in the meta-linguistic awareness they have of even the local regional accent. Seventy-one percent of subjects could identify the regional accent of their hometown. Although that is a majority of subjects, it still leaves a third of the subjects who cannot correctly identify the local accent. This would be the most frequently encountered regional accent of daily interactions for many subjects, yet not all of them could correctly identify it. The situation is even worse for a non-familiar regional accent; only 43% of all subjects could state that it was, at the very least, not a local regional accent. As predicted, subjects are better at interpreting a familiar accent, and identifying what it represents, but are less skilled at doing the same for an unfamiliar regional accent. Similarly, subjects appear to use recognition of the local speakers in the two experimental tasks to discriminate between accents, but are not able to do the same for the non-local speakers in most cases. These findings support the hypothesis that more frequently encountered kinds of variation are more readily interpretable to children.

The second hypothesis presented by Foulkes and Docherty (2006) about how acquisition of social categories occurs is that the more apparent the social category, the easier it will be to link it to a characteristic kind of speech, or exemplar category for that variety. In the United States, regional differences aren't generally associated with differences in appearance. This would make identifying regional variation much more difficult than age-based or some categories of ethnicity-based variation. So it is particularly interesting to look at the results of Task 2, the Similarity Judgment Task. In particular, the results of the African-American subjects are important, to see whether children interpret the variation in the task to be about race or regional accents, and whether they find any similarities between themselves and either set of speakers.

The African-American subjects choose the local speakers as sounding similar to themselves at the same rate as the Caucasian sub-group, and show the same pattern of correlations with identifying the local regional accent and being able to say how regional accents are distributed (Q3 and Q5 on the Awareness Task) as the entire subject group. I had not expected the African-American sub-group to do this; I had expected that because the local regional accent is associated with Caucasians in that community, the African-American subjects would not judge the local accent to be similar to their own speech. In short, I expected ethnic variation to be more salient than regional variation to those subjects, in part because of its

transparent link to another marker of that social category and the experience they have with ethnic variation on a daily basis.

This is not to say that five to seven year old children don't understand ethnic variation. What it may mean is that children don't interpret variation in the same way as adults do. Clearly the African-American subjects accurately identify the local speakers, meaning that particular variety is linked to a category for local individuals. I don't know from these experiments how broad that category is, or whether it includes all the ethnolects of their hometown. But unlike the Outsider subjects, they are as accurate identifying the local speakers and as likely to use that information in discrimination as the Caucasian subjects. There is no confusion about who is a local speaker for the African-American participants in this study.

This suggests at least two explanations for their willingness to judge a local speaker similar to themselves. The first possibility is that ethnic variation isn't salient to them, or isn't interpreted as adults, myself included, might expect. Perhaps race is not a defining basis on which to judge differences in speech, and they actually classify all of the varieties of their hometown as local, but don't further sub-divide them based on ethnicity.

The second possibility is that they don't consider ethnicity to be more important than shared place of residence when defining similarity between speakers. Although the African-American children may have also been able to identify the speakers as Caucasian (although they were not asked to do so), they thought that the difference in race did not outweigh the fact that the speakers were from the same town as they, and therefore were willing to claim those speakers sounded similar to themselves.

Although this nominally was a forced-choice task, I left open the possibility for children to respond that neither speaker was similar, or not to respond at all. I specifically did this to see if the African-American participants would prefer no response over choosing either speaker as similar, since all speakers were Caucasian. I predicted at the very least that African-American children would use some other property of speech to base similarity on, or would comment about the fact that none of the speakers sounded like them. None of these things happened; they systematically chose the local speakers as sounding similar.

At first glance, this finding seems to refute what Foulkes and Docherty (2006) predict, that a visible kind of social variation is more easily acquired than a less apparent one. But I don't know from this finding whether they can interpret ethnic variation, just that it doesn't

override other kinds of variation in their interpretation of social categories marked in speech. At this point in development, they may just give equal weight to all the different kinds of social variation they perceive. The fact the speakers were of a different race did not cancel out the fact that they were from the same town as the subjects. There is a possibility, however, that African-American children may not yet be attuned to racial differences. Aboud and Amato (2001) discuss social-psychology experiments whose results show that only starting around age five do children perceive race as a distinguishing characteristic of individuals; the salience of race to children has to be learned. If this is the case for children in this context as well, then it may be that the subjects were just starting to grasp that this is a social dimension to which adults assign importance. Thus, they may not have acquired the social background that would cause them to weight ethnic variation any differently than regional variation in finding speaker similarity. That is to say that children's attention will be re-focused on certain kinds of variation as they learn more about the social structure of their community, and this will affect how they perceive speech, including their own.

Foulkes and Docherty (2006) refer to a stage in acquisition where children shift their attention from statistical weighting of exemplars to social weighting; meaning that they would be able to interpret social variation with which they have less experience because of an increased understanding of what that kind of variation represents. Foulkes and Docherty specifically reference a Kerswill and Williams study (2000) in which children, upon entering school, suddenly shift from using the regional variety of their parents in speech to those of the local regional variety, as used by peers in school. This sudden shift to using a variety they have only just encountered is indicative of a shift in focus from the more frequent variety (the home variety) to one of social importance (that of their peers). I suspect this is in part because once they enter school, children become aware of the age-based organization imposed on their social world; they express this by adopting the appropriate kinds of variation. Similarly as other kinds of social differences are given emphasis, they likely weight those over other kinds of variation they hear. Ethnicity may be one of these, but further studies will be needed to determine this definitively.

Finally, it should be mentioned that although there is a common local variety of U.S. English spoken in the town where this study was conducted, this is not the only variety heard in the town. Standard American English, African-American English, Hispanic English and English

from L2 speakers are all heard in this town as well. Yet children of speakers of these other varieties don't seem to suffer from confusion when identifying the local regional accent as Outsiders do. Future work will explore what the boundaries are of children's representation of the local accent are, how inclusive it is of other ethnolects and dialects, and what kinds of exposure result in different definitions of local. I expect the definition of local to vary on an individual basis, and more detailed work looking at the social connections between subjects and speakers with other accents may help explain how exemplar representations or social categories of speakers develop.

6.2.2 Development of Exemplar Categories based on Interaction with Specific Individuals

The other hypothesis from the literature on the development of exemplar categories for social variation in childhood can be discussed in light of the findings from this study. Munson (2010) predicts that social categories are linked to kinds of speech based on interaction with specific, identifiable individuals in the children's lives. This is evidenced not only in the superior performance of all subjects in the current study identifying and interpreting the local accents over non-local accents, but also in the Outsiders' performance on two specific measures. First, many Outsiders subjects have difficulty recognizing the local accent in the Awareness Task, more so than most other sub-groups except the Bilinguals. Secondly, they apparently do not use identification of the local speakers to assist in making matches in the experimental tasks, based on the results of the statistical analyses. I surmise this is because they don't yet understand, based on their input, that there is a single, local variety. They hear at least two different regional varieties in their input, both from speakers who are all local from the child's point of view. This interaction with two speakers who have different regional accents but live in the same place makes it difficult for them to construct a local accent category. Children for whom the input is unambiguous, where there is one regional accent associated with the hometown, have an easier time making the association. Were children not dependent on their interactions with specific, identifiable individuals to build a category for local speech, the Outsiders would have a representation similar to the Insiders for local speech. All of the subjects live in the same community and attend the same schools, so with the exception of the home, have similar experiences hearing the local accent in the community. However, there is evidence in the results of this study that children differ in their interpretations of regional accent based on their

parents' backgrounds. I suspect that once Outsider children understand that one parent is not from the hometown, the local regional accent is disambiguated and they have a similar representation as the Insiders for the local regional accent. This could be tested in future studies, to see if Outsider children who are aware of where both parents are from perform differently than those who do not.

Surprisingly, this confusion is not at all mediated by travel for the Outsiders. I initially had thought that traveling to another region and hearing the variety of the parents spoken in another location would help clarify the link between locality and speech variety for Outsider children. This appeared not to be the case. However, travel was found to increase likelihood of identifying the local speaker in the Similarity Judgment Task for the Insider children, whose input is mainly in the local regional accent. Perhaps for the Insiders, since they have a well-established category for what a local accent sounds like, knowledge of another regional accent only serves to strengthen that category by providing a contrast. The link between a particular variety and location is clear for Insiders, whereas the Outsiders are receiving input that does not suggest a single variety exists locally.

Despite the confusion of the Outsiders when explicitly drawing on knowledge of how regional accent marks local speakers, it must be kept in mind that the Outsiders were no worse than the subject group as a whole, or than the Insiders, in discriminating between regional accents. This was true for both the ABX task and the Similarity Judgment Task. The differences between the sub-groups came when they were trying to use social knowledge about regional accents to help with discrimination (no correlations were found between this knowledge and the results of either discrimination task for Outsiders) or state where a speaker with the local regional accent was from.

This leads to the question of how crucial exemplar categories are to discrimination. If the category for the local regional accent is still emergent for Outsider children, how is it that they discriminate with accuracy equal to children who have a clear understanding of the local accent? I suspect the answer is that the linguistic exemplar categories are equally robust for both Insider and Outsider children. The Outsiders live in the same community and attend the same schools as the Insiders, and many Insiders have at least one parent from outside the local community. Even if absolute levels of exposure are the same, they all have experience hearing the local regional accent regularly, and as a result, should have an exemplar representation of it.

Beckman et al (2007) propose a model in which there are emergent categories of social knowledge, referred to by Munson (2010) as a "lexicon of talkers" that generalizes to create connections between linguistic exemplars and social categories of speakers. Children learn to connect exemplar categories of speech to their social indices through experience with individual talkers whose social category memberships are known, and thus can be associated with the characteristic speech for those groups. At this stage, children are thought to be highly dependent on associating a speech variety with a particular individual; the social categories are not well generalized and are linked to a few known speakers with whom they interact. This would explain why Outsiders showed no evidence of using recognition of the local speakers in either experimental task. They do not have an association with just one regional speech variety and a small set of local speakers. Instead, they have a small set of (incorrectly identified) local speakers – the parents - associated with at least two regional varieties.

Thus, when the input fails to activate the linguistic category from one of the core (presumed) local speakers in the child's lexicon of talkers, the association of that speech variety with the local area fails. The input did not match the speech exemplars of all of the local individuals, and thus did not activate the local social category, which in turn does not allow the child to recognize and use the local category to help find matches in the experimental tasks. The social category must be linked to the correct set of linguistic exemplars before it can be utilized in discrimination.

Again, recall that Outsiders did no worse than any other group overall in discrimination; the positive correlation between identifying the local accent and matching accents reflects the probability of a match in a particular trial, not overall performance on either task. The lack of significant differences in performance between Insiders and Outsiders, or Outsiders and the whole subject group, indicates that they were able to make matches based on the linguistic exemplar categories they have for the local accent, when they were able to reason through the differences in accent using overt knowledge of the source of regional accents. That is to say, Outsiders who recognize that one of those categories is not actually local can use this information and match the local accents/choose the local accent as most similar to themselves. This points to the importance of having socio-indexical information correctly associated with a linguistic category of social variation; those Outsiders who label both of the accents they hear in

their environment as local can't use socio-indexical information as a basis for matching speakers and accents as other subjects appear to do.

6.2.3 The Role of Socio-Indexical Information

In the introduction, I addressed a study by Drager (2010), in which high school students were asked to identify which of their classmates ate lunch in the Common Room in school, based on the variant of the word *like* that individual used. The subjects were inaccurate when using the phonetic realization of *like* to identify where the speaker ate lunch. They were instead using the realization of *like* as a means to assess whether the speaker was "normal" or not, based stereotypes of how "normal" girls would pronounce that word. They also attempted to identify which classmate produced each token of *like* in order to complete the task. I surmised that because the two social groups Drager identified through ethnographic study, the Common Room Girls and the Non-Common Room Girls weren't salient social categories to the subjects, *like* could not index those two groups. For that reason, the subjects could not rely on the socio-indices of *like* to help them identify the correct social group. Put another way, the social categories Drager wanted them to associate *like* within the experiment weren't categories the subjects were aware of, and had no socio-indices for.

This study was relevant to my own, as I thought it likely a similar situation would exist for the subjects of my study; either they would attribute the accents to individual or some other kind of variation, and/or they would not be aware of regional variation at all.

I suspected a lack of overt awareness of regional variation would negatively impact the ability to identify it. Developing a category based on the statistical regularities in the input is possible in the absence of a label or socio-indexical information. However, it seriously limits the ability to identify speakers if that category isn't associated with a particular group, or the group is unnamed. It makes reporting having heard an accent next to impossible, since the listener may not know what the source of variation even is.

I found that for the local accent, most were aware of the category the local accent indexed, and were quite accurate when using socio-indexical information to help them discriminate speakers in both experimental tasks. However, for the unfamiliar accent, which indexed nothing for most subjects and likely was not associated with any kind of social category, the accuracy rate was worse in the ABX task, and it appears that they attempted to use

stereotypes or some kind of general knowledge of accents to help them, which further hurt their accuracy in discrimination.

Taken together, I believe these two studies indicate that although listeners may be subconsciously extracting statistical patterns of linguistic variables in their input, and associating those with particular categories, that they cannot operationalize this information and use it to identify groups until social indexical information rises above their level of consciousness. That is to say they are not storing tokens to find statistical regularities in speech, and associate those with social categories. I just don't believe that they can use that information to identify a social group until they are consciously aware of the groups' existence. In other words, listeners may not be able to discover new social groups through experience, but may need to explicitly learn about them and label them in order to begin associating experiential tokens with the correct categories.

What a category looks like up to the point of conscious awareness is unclear. Based on Drager's finding that subjects attempted to identify the speaker rather than the lunch group in her task, I would expect that it is a collection of individuals, a lexicon of speakers, as described by Beckman et al. (2007), rather than a social category. Perhaps once a category is labeled, the accumulated tokens can be correctly sorted into that category; that remains to be tested. It also appears that in the absence of a social index, that identification of speakers based on speech samples is subject to stereotypes and "declarative knowledge" about the group, which can skew identification as much as lack of exposure to the linguistic variable(s) in question.

Further work is needed to understand not only how linguistic categories may form, but also what their relationship to social categories is.

CHAPTER 7

Conclusion

This study asked how children aged five to seven perceive regional accents, and what their abilities say about the representations they have of this particular kind of social variation.

The study asked four main questions, each assessed with its own instrument.

First: Can children discriminate between regional accents, when no reference to region or accent is made in the task (i.e. is this kind of variation salient to them?)

Second: Can they identify a speaker using information indexed by a regional accent?

Third: What kinds and levels of exposures to accents, regional and other, affect ability to report hearing different kinds of variation?

Four: What is their state of metalinguistic awareness about regional accents, and does it have an effect on their ability to report hearing this kind of variation?

Below I summarize the findings of each task, answering the questions they presented.

The results of the ABX discrimination task clearly show that even in the absence (or perhaps because of the absence) of references to accents and regions, children discriminate between a familiar, local and non-familiar, non-local regional accent. Monolinguals performed with the highest rate of accuracy, but even with bilingual subjects included, performance of the subject group was above chance.

As to the second question, it also appears that most children aged five to seven are aware that the local regional accent marks a person from their hometown, although hearing two different regional accents at home complicates identification of a local regional accent.

Interestingly, speaking an ethnolect did not have the same complicating effect in identifying the local regional accent that having parents from outside the area did. I attribute this

to the transparency of ethnic variation to children, and the relative difficulty of identifying the source of regional variation.

Children will also use the local regional accent as a marker of similarity, claiming that they sound like other people from their hometown. This is true even in cases where the child speaks with another kind of accent or ethnolect (such as AAE, or Hispanic English). I believe that children's willingness to identify with the local accent demonstrates that they associate at least basic socio-indexical information about where a speaker is from with the linguistic features characterizing an accent. If they were not aware that an accent indexes the speaker's place of residence, a number of children in this study would have no basis to claim similarity with the local speakers, which they did 70% of the time in this study.

Question three, addressed by the Awareness Task, shows that children can accurately identify local speakers based on accent, but were less confident stating that a non-local speaker was not from their hometown, remaining agnostic about where the non-local speaker was from most of the time. This seems to suggest that children are most confident in interpreting the indices of regional accents with which they have intensive experience, as opposed to those that are less familiar. They also proved to be unable to extrapolate on those indices most of the time, to indicate that the difference between the two sets of speakers heard in this study was their place of origin. This may suggest that socio-indexical information isn't available for reasoning at this stage; that is to say they cannot conclude that a speaker who does not sound like local speakers has another regional accent.

Question four asked whether experience hearing regional accents improved the ability to discriminate between them. No specific kinds of experience with non-local regional accents measured in the parent questionnaire were found to influence performance on either experimental task. At face value, this would seem to contradict the claim that categories or exemplar clouds for accents are based on experience hearing them. However, at the ages of five to seven, the experience reported may be of insufficient quantity to have established a meaningful exemplar category for another regional accent. Alternately, the child may have an excellent representation of the linguistic qualities of a particular regional accent, but may not have the requisite social knowledge to recognize or label it as regional variation. Support for this second interpretation of why the measures of exposure from the Parent Questionnaire do not show a correlation with the experimental tasks was found in examining sub-group results. I

divided subjects up into multiple sub-groups in an attempt to capture in a more general way the exposure they may have had to other kinds of linguistic variation.

One of those groups was the Outsiders, or children with at least one parent from outside the region. For this group, unlike the other sub-groups, there were no correlations between their experimental task responses and their awareness of the local accent. They were, as a group, also less accurate than all of the other sub-groups to have identified the local speakers correctly in the Awareness Task. The explanation for this result is that the Outsiders had not yet sorted out what the local accent is because of the exposure they've had to two regional accents in the home. This is a source of confusion, or the distinction is unimportant to them, since they have daily contact with two people who live in their town who possibly speak with two very different regional varieties. This makes it difficult for them to characterize what a local accent is at first, until they understand that one of the parents is not from the town originally, allowing them to isolate the local accent.

It is interesting to note that this was not the case for children who have experience hearing ethnic variation at home, perhaps because the "source" of that variation is transparent. In contrast, regional variation in the United States may seem completely arbitrary to children at this age, since there are no visible means of differentiating individuals from different areas of the country.

This study shows that five to seven year old children are able to discriminate between regional accents. It also suggests that their understanding of regional accent is dependent on direct experience with speakers of that accent, and possibly with a limited set of familiar individuals. Furthermore, most children seem able to correctly identify the social characteristics of a speaker based on regional accent, although it is still unclear how adult-like this social categorical information is, and how individually variable it is.

This study had predicted that general meta-linguistic awareness of regional accents, and exposure to those accents, would improve the ability to discriminate. The correlation between these two factors and discrimination is not direct; an increase in either of these factors does not directly predict an improvement in overall discrimination ability. However, the meta-linguistic awareness and exposure to regional accents may determine how the child approaches the task, and what information he or she uses to discriminate between the two regional accents.

Finally, the results of this study match well with predictions made about how exemplar categories would develop in children, based on experience and exposure to individuals with particular regional accents. Further study on the development of the social categories would help to further illuminate how children acquire the ability to interpret sociolinguistic variation.

In sum, I find that children can discriminate between a familiar and an unfamiliar regional accent, and that they have the ability to interpret the familiar accent for socio-indexical information about the speaker. However, it appears that the creation of a linguistic or social exemplar representation for a familiar accent/social category does not at all assist in interpreting unfamiliar or non-local regional accents. It appears that children remain agnostic about what kind of variation an unfamiliar regional accent represents when they don't have an established category for that other regional accent.

APPENDICES

APPENDIX 1

Parent Questionnaire

<table>
<tr><td colspan="2">Child's Gender: M F
Child's Birthday:</td></tr>
<tr><td colspan="2">Localization</td></tr>
<tr><td>Birth year:</td><td>Place of birth:</td></tr>
<tr><td rowspan="2">Father's birthplace and current location</td><td>Father of father birthplace and current location</td></tr>
<tr><td>Mother of father birthplace and current location</td></tr>
<tr><td rowspan="2">Mother's birthplace and current location</td><td>Father of mother birthplace and current location</td></tr>
<tr><td>Mother of mother birthplace and current location</td></tr>
<tr><td colspan="2">Other places of residence</td></tr>
<tr><td colspan="2">Other places child lived since birth and how long:</td></tr>
<tr><td colspan="2">Other places father lived and how long:</td></tr>
<tr><td colspan="2">Other places mother lived and how long:</td></tr>
<tr><td colspan="2">Contact with others outside of region</td></tr>
<tr><td colspan="2">Has your child ever travelled outside of your home region (100 miles or more)? Yes No

If yes, where to (list all relevant places)?</td></tr>
</table>

How often and for how long? *Example: Every summer, 2 weeks visiting cousins in Mobile, Alabama*

Does your child interact regularly with people who talk differently than most people in your home area (i.e. people you consider to have an accent, either foreign or different US accent)? Yes No

If yes, what relationship do they have to the child (friend, cousin, babysitter, etc)? List as many as possible people with accents that the child has regular contact with.

How often does the child speak to that person/people?

Does your child watch TV programs from outside the U.S. (for example, British TV programs/movies)? Yes No
If Yes, please give titles:

Does your child speak any other languages? Yes No

If yes, which ones, and how well does the child speak them?
Example: Spanish, studied in school for 1 year

Does the child have any relatives who are non-native English speakers, or who speak a language other than English to the child? Yes No

If yes, what is their relationship to the child, and what language do they speak?

How often does the child speak to that person?

Has your child ever commented on or tried to imitate someone else's accent? Please give details, if you can (use back if necessary):

Has your child ever commented on his or her own way of speaking, or on someone else's way of speaking? Please give details, if you can (use back if necessary):

APPENDIX 2

List of Stimuli

Word	Class
Day	FACE
Hay	FACE
Neigh	FACE
Play	FACE
Buy	PRICE
Light	PRICE
Night	PRICE
Pie	PRICE
Tie	PRICE
Boat	GOAT
Coat	GOAT
Goat	GOAT
Note	GOAT
Heel	peel
Peel	peel

Word	Class
Seal	peel
Wheel	peel
Sail	tail
Sell	tail
Tail	tail
Well	tail
Juice	GOOSE
News	GOOSE
Tune	GOOSE
You	GOOSE
Bad	Filler
Daughter	Filler
Farmers	Filler
Forest	Filler
Little	Filler

APPENDIX 3

Output of Statistical Models

Appendix 3.1 Awareness Task (Chapter 3)

```
> dat=read.table("awarenessglm.txt", T)
> mod=lm(Metascore ~ Age + Outsider, data=dat)
> summary(mod)

Call:
lm(formula = Metascore ~ Age + Outsider, data = dat)

Residuals:
   Min     1Q  Median     3Q    Max
-3.4511 -1.1898 -0.0149  1.3202  2.4412

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.32218    3.41598  -1.851  0.06889 .
Age          0.13031    0.04821   2.703  0.00883 **
Outsider    -0.37100    0.21972  -1.689  0.09626 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.473 on 63 degrees of freedom
Multiple R-squared: 0.1458,   Adjusted R-squared: 0.1187
F-statistic: 5.376 on 2 and 63 DF,  p-value: 0.006985

> mod=lm(Q5 ~ Age + Outsider, data=dat)
Error in eval(expr, envir, enclos) : object 'Q5' not found
> mod=lm(Q3 ~ Age + Outsider, data=dat)
> summary(mod)

Call:
lm(formula = Q3 ~ Age + Outsider, data = dat)

Residuals:
    Min      1Q   Median      3Q     Max
```

```
-0.71025 -0.29768 -0.09917  0.39095  0.90083

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.20913    0.95432  -3.363 0.001317 **
Age          0.05157    0.01347   3.829 0.000299 ***
Outsider    -0.14694    0.06138  -2.394 0.019664 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4114 on 63 degrees of freedom
Multiple R-squared: 0.2552,   Adjusted R-squared: 0.2316
F-statistic: 10.79 on 2 and 63 DF,  p-value: 9.31e-05

> mod=lm(Q2B ~ Age + Outsider, data=dat)
> summary(mod)

Call:
lm(formula = Q2B ~ Age + Outsider, data = dat)

Residuals:
   Min      1Q  Median      3Q     Max
-0.6535 -0.4375 -0.1330  0.4528  0.8670

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.69238    1.10423  -0.627   0.5329
Age          0.01771    0.01559   1.136   0.2602
Outsider    -0.18056    0.07103  -2.542   0.0135 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4761 on 63 degrees of freedom
Multiple R-squared: 0.1144,   Adjusted R-squared: 0.08626
F-statistic: 4.068 on 2 and 63 DF,  p-value: 0.02179

> mod=lm(Q2A ~ Age + Outsider, data=dat)
> summary(mod)

Call:
lm(formula = Q2A ~ Age + Outsider, data = dat)

Residuals:
   Min      1Q  Median      3Q     Max
-0.8439 -0.4158  0.1759  0.2822  0.6707
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.102413   1.053142   0.097  0.92284
Age          0.009886   0.014865   0.665  0.50843
Outsider    -0.202895   0.067740  -2.995  0.00392 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.454 on 63 degrees of freedom
Multiple R-squared: 0.1333,   Adjusted R-squared: 0.1058
F-statistic: 4.846 on 2 and 63 DF,  p-value: 0.01102

> mod=lm(Q1A ~ Age + Outsider, data=dat)
> summary(mod)

Call:
lm(formula = Q1A ~ Age + Outsider, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8492 -0.5062  0.1728  0.3097  0.6907

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.47018    1.02055  -1.441   0.1547
Age          0.02917    0.01440   2.025   0.0471 *
Outsider     0.15328    0.06564   2.335   0.0227 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.44 on 63 degrees of freedom
Multiple R-squared: 0.1251,   Adjusted R-squared: 0.09734
F-statistic: 4.505 on 2 and 63 DF,  p-value: 0.01484

> mod=lm(Q1B ~ Age + Outsider, data=dat)
> summary(mod)

Call:
lm(formula = Q1B ~ Age + Outsider, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59423 -0.48615  0.02711  0.48609  0.60816

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -0.749896   1.176531  -0.637    0.526
Age          0.017686   0.016606   1.065    0.291
Outsider     0.004921   0.075677   0.065    0.948

Residual standard error: 0.5072 on 63 degrees of freedom
Multiple R-squared: 0.01769, Adjusted R-squared: -0.0135
F-statistic: 0.5671 on 2 and 63 DF,  p-value: 0.57
```

Appendix 3.2 ABX Task (Chapter 4)

ALL SUBJECTS

```
dat=read.table("Task_A_no_filler.txt", T)
> dat.lmer = lmer(Answer ~   Q3 + Q5 + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q5 + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1981 2007 -985.3     1971
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.024498 0.15652
 Word    (Intercept) 0.220101 0.46915
Number of obs: 1516, groups: Subject, 61; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.4688     0.1343   3.491 0.000481 ***
Q3yes         0.2244     0.1268   1.769 0.076839 .
Q5yes        -0.2685     0.1284  -2.091 0.036556 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr) Q3yes
Q3yes -0.475
Q5yes  -0.134 -0.346
```

MONOLINGUALS

```
dat=read.table("Task_A_Monolonguals.txt", T)
> dat.lmer = lmer(Answer ~   Q3 + Q5 + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q5 + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1602 1628 -796.2     1592
Random effects:
```

```
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.00000  0.00000
 Word    (Intercept) 0.31011  0.55688
Number of obs: 1244, groups: Subject, 50; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.5004     0.1597   3.134  0.00172 **
Q3yes         0.2754     0.1380   1.996  0.04590 *
Q5yes        -0.3388     0.1270  -2.667  0.00766 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr) Q3yes
Q3yes -0.508
Q5yes  -0.173 -0.287
```

BILINGUALS

```
> dat=read.table("Task_A_bilinguals.txt", T)


> dat.lmer = lmer(Answer ~  Q3 + Q4 + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q4 + (1 | Word) + (1 | Subject)
   Data: dat
   AIC   BIC logLik deviance
 372.8 390.8 -181.4    362.8
Random effects:
 Groups  Name        Variance   Std.Dev.
 Word    (Intercept) 8.6340e-12 2.9384e-06
 Subject (Intercept) 1.8773e-01 4.3328e-01
Number of obs: 272, groups: Word, 25; Subject, 11

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.4500     0.2133   2.110   0.0349 *
Q3yes        -0.3682     0.4685  -0.786   0.4319
Q4yes         0.5062     0.7319   0.692   0.4892
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
      (Intr) Q3yes
Q3yes -0.455
Q4yes  0.000 -0.507
```

CAUCASIANS

```
dat=read.table("Task_A_C.txt", T)
dat.lmer = lmer(Answer ~  Q3 + Q5 + (1|Word) + (1|Subject), family="binomial",
data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q5 + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1234 1258   -612     1224
Random effects:
 Groups  Name        Variance   Std.Dev.
 Subject (Intercept) 9.7161e-13 9.8570e-07
 Word    (Intercept) 2.2010e-01 4.6914e-01
Number of obs: 946, groups: Subject, 38; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.5779     0.1602   3.607  0.00031 ***
Q3yes         0.1837     0.1525   1.204  0.22849
Q5yes        -0.3709     0.1417  -2.617  0.00887 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr) Q3yes
Q3yes -0.572
Q5yes  -0.244 -0.222
```

AFRICAN-AMERICANS

```
> dat=read.table("Task_A_AA.txt", T)
> dat.lmer = lmer(Answer ~   Q3 + Q5 + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q5 + (1 | Word) + (1 | Subject)
   Data: dat
   AIC   BIC logLik deviance
 395.1 413.6 -192.5    385.1
Random effects:
```

```
 Groups  Name        Variance Std.Dev.
 Word    (Intercept) 0.47314  0.68785
 Subject (Intercept) 0.00000  0.00000
Number of obs: 298, groups: Word, 25; Subject, 12

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.2250     0.2506   0.898   0.3692
Q3yes         0.6076     0.3049   1.993   0.0463 *
Q5yes        -0.3856     0.3055  -1.262   0.2069
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr) Q3yes
Q3yes -0.573
Q5yes  0.000 -0.527
>
```

INSIDERS

```
> dat=read.table("Task_A_Insiders.txt", T)
> dat.lmer = lmer(Answer ~   Q3 + Q5 + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q5 + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1302 1327   -646     1292
Random effects:
 Groups  Name        Variance   Std.Dev.
 Subject (Intercept) 2.4649e-11 4.9648e-06
 Word    (Intercept) 2.3988e-01 4.8977e-01
Number of obs: 996, groups: Subject, 40; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.4043     0.1602   2.524   0.0116 *
Q3yes         0.3491     0.1624   2.150   0.0315 *
Q5yes        -0.3352     0.1482  -2.261   0.0237 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr) Q3yes
```

```
Q3yes -0.573
Q5yes  -0.089 -0.405
>
```

OUTSIDERS

```
dat=read.table("0_NTOWN_PARENT.txt", T)
dat.lmer = lmer(Answer ~  Q3 +  Q5 + (1|Word) + (1|Subject),
family="binomial", data=dat)
dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q3 + Q5 + (1 | Word) + (1 | Subject)
 Data: dat
 AIC   BIC logLik deviance
 397.8 416.3 -193.9    387.8
Random effects:
Groups  Name        Variance Std.Dev.
Word    (Intercept) 0.49188  0.70134
Subject (Intercept) 0.00000  0.00000
Number of obs: 299, groups: Word, 25; Subject, 12

Fixed effects:
 Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.3026     0.2535   1.193   0.2328
Q3            0.4319     0.2479   1.742   0.0815 .
Q5           -0.1850     0.2486  -0.744   0.4567
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

INTERSTIMULUS INTERVAL ANALYSIS:

```
> dat.lmer = lmer(Answer ~  ISI + (1|Word) + (1|Subject), family="binomial",
data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ ISI + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 2266 2288  -1129     2258
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.087248 0.29538
 Word    (Intercept) 0.347806 0.58975
Number of obs: 1821, groups: Subject, 61; Word, 30

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   0.8046     0.1349   5.964 2.46e-09 ***
ISIshort     -0.2092     0.1075  -1.946   0.0517 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 3.3: Task 2 (Chapter 5)

ALL SUBJECTS

```
> dat=read.table("TaskB_no_filler.txt", T)

dat=read.table("TaskB_no_filler.txt", T)
> dat.lmer = lmer(Answer ~  Q2A + Q2B + Q3 + Imitate + Comment + (1|Word)
+ (1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Comment + (1 | Word) + (1 |
Subject)
   Data: dat
  AIC  BIC logLik deviance
 1711 1753 -847.5     1695
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.54097  0.73551
 Word    (Intercept) 0.13638  0.36930
Number of obs: 1502, groups: Subject, 61; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.4629     0.2344   1.975   0.0483 *
Q2Ayes        0.6468     0.2693   2.402   0.0163 *
Q2Byes       -0.5139     0.3003  -1.712   0.0870 .
Q3yes         0.7473     0.2915   2.564   0.0104 *
Imitate       0.5121     0.2583   1.983   0.0474 *
Comment      -0.2467     0.2611  -0.945   0.3446
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Q2Ayes Q2Byes Q3yes  Imitat
Q2Ayes  -0.368
Q2Byes  -0.171 -0.364
Q3yes   -0.075 -0.064 -0.495
Imitate -0.152 -0.094 -0.136  0.123
Comment -0.343 -0.148  0.255 -0.114 -0.443
>

> dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
```

```
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1673 1700 -831.7     1663
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.63675  0.79797
 Word    (Intercept) 0.14467  0.38035
Number of obs: 1477, groups: Subject, 60; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.25044    0.31045   0.807  0.41984
Awareness    0.22636    0.07595   2.980  0.00288 **
Exposure     0.08708    0.09017   0.966  0.33419
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) Awrnss
Awareness -0.554
Exposure  -0.641 -0.091
```

AFRICAN-AMERICANS

```
> dat=read.table("TaskB_AA.txt", T)
> dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
   Data: dat
   AIC   BIC logLik deviance
 386.8 405.7 -188.4    376.8
Random effects:
 Groups  Name        Variance Std.Dev.
 Word    (Intercept) 0.102729 0.32051
 Subject (Intercept) 0.093713 0.30613
Number of obs: 321, groups: Word, 25; Subject, 13

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08344    0.32991  -0.253 0.800336
Awareness    0.30267    0.08399   3.604 0.000314 ***
Exposure     0.12465    0.12036   1.036 0.300365
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) Awrnss
Awareness -0.400
Exposure  -0.712 -0.123


dat=read.table("Task_B_AA.txt", T)
> dat.lmer = lmer(Answer ~  Q2A + Q2B + Q3 + Imitate + Comment + (1|Word)
+ (1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Comment + (1 | Word) + (1 |
Subject)
   Data: dat
   AIC BIC logLik deviance
 426.3 457 -205.1    410.3
Random effects:
 Groups  Name        Variance Std.Dev.
 Word    (Intercept) 0.035046 0.18721
 Subject (Intercept) 0.029767 0.17253
Number of obs: 346, groups: Word, 25; Subject, 14

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4676     0.2848  -1.642  0.10064
Q2Ayes        0.8949     0.2872   3.116  0.00184 **
Q2Byes       -0.8029     0.5919  -1.357  0.17495
Q3yes         1.4192     0.6616   2.145  0.03195 *
Imitate      -0.2249     0.3645  -0.617  0.53723
Comment       1.1403     0.3608   3.160  0.00157 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
          (Intr) Q2Ayes Q2Byes Q3yes  Imitat
Q2Ayes    -0.588
Q2Byes     0.147 -0.358
Q3yes     -0.296  0.261 -0.884
Imitate    0.028  0.001  0.358 -0.530
Comment   -0.550  0.255 -0.469  0.578 -0.610
>
```

CAUCASIANS

```
("Task_B_mono_CC.txt", T)
dat.lmer = lmer(Answer ~  Q2A + Q2B + Q3 + Imitate + Comment + (1|Word) + (1|Subject),
family="binomial", data=dat)
dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Comment + (1 | Word) + (1 |     Subject)
  Data: dat
  AIC  BIC logLik deviance
 963.7 1002 -473.9    947.7
Random effects:
 Groups  Name       Variance Std.Dev.
 Subject (Intercept) 0.47496  0.68918
 Word    (Intercept) 0.13149  0.36262
Number of obs: 884, groups: Subject, 36; Word, 25

Fixed effects:
         Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.0975     0.3540   3.100  0.00193 **
Q2Ayes        0.4035     0.4254   0.948  0.34289
Q2Byes       -0.6531     0.3804  -1.717  0.08596 .
Q3yes         0.7910     0.3348   2.363  0.01814 *
Imitate       0.5801     0.3181   1.824  0.06818 .
Comment      -0.8898     0.3282  -2.711  0.00671 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr) Q2Ayes Q2Byes Q3yes  Imitat
Q2Ayes  -0.395
Q2Byes  -0.119 -0.523
Q3yes   -0.170 -0.139 -0.286
Imitate -0.229 -0.165 -0.027  0.282
Comment -0.363 -0.215  0.338 -0.093 -0.305
> dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
  Data: dat
 AIC BIC logLik deviance
 968 992   -479      958
Random effects:
 Groups  Name       Variance Std.Dev.
 Subject (Intercept) 0.72870  0.85364
 Word    (Intercept) 0.13159  0.36276
```

Number of obs: 884, groups: Subject, 36; Word, 25

Fixed effects:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.97073 0.52472 1.850 0.0643 .
Awareness 0.09948 0.11998 0.829 0.4070
Exposure -0.02824 0.12065 -0.234 0.8149

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
(Intr) Awrnss
Awareness -0.735
Exposure -0.643 0.082

MONOLINGUAL OUTSIDERS

dat.lmer = lmer(Answer ~ Awareness + Exposure + (1|Word) + (1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
Data: dat
AIC BIC logLik deviance
297.2 314.7 -143.6 287.2
Random effects:
Groups Name Variance Std.Dev.
Word (Intercept) 0.0000 0.00000
Subject (Intercept) 0.7142 0.84511
Number of obs: 246, groups: Word, 25; Subject, 10

Fixed effects:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 2.9828 1.6417 1.817 0.0692 .
Awareness -0.2415 0.3217 -0.751 0.4529
Exposure -0.3591 0.3053 -1.176 0.2394

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
(Intr) Awrnss
Awareness -0.655
Exposure -0.774 0.067

> dat.lmer = lmer(Answer ~ Q2A + Q2B + Q3 + Imitate + Travel + (1|Word) + (1|Subject), family="binomial", data=dat)

```
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Travel + (1 | Word) + (1 |
Subject)
   Data: dat
   AIC   BIC logLik deviance
 285.5 313.6 -134.8    269.5
Random effects:
 Groups  Name        Variance Std.Dev.
 Word    (Intercept)  0        0
 Subject (Intercept)  0        0
Number of obs: 246, groups: Word, 25; Subject, 10

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.63901    1.30857   2.781  0.00542 **
Q2Ayes       0.49931    0.34813   1.434  0.15149
Q2Byes      -1.95048    0.41097  -4.746 2.07e-06 ***
Q3yes        0.93704    0.42152   2.223  0.02622 *
Imitate      0.05317    0.39285   0.135  0.89234
Travel      -3.26204    1.12331  -2.904  0.00368 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Q2Ayes Q2Byes Q3yes  Imitat
Q2Ayes  -0.207
Q2Byes  -0.123 -0.426
Q3yes   -0.442  0.137 -0.269
Imitate -0.518  0.104  0.031  0.559
Travel  -0.946  0.043  0.162  0.273  0.290


 dat.lmer = lmer(Answer ~  Q2A + Q2B + Q3 + Imitate + Comment + (1|Word) +
(1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Comment + (1 | Word) + (1 |
Subject)
   Data: dat
   AIC   BIC logLik deviance
 285.5 313.6 -134.8    269.5
Random effects:
 Groups  Name        Variance Std.Dev.
 Word    (Intercept)  0        0
 Subject (Intercept)  0        0
```

```
Number of obs: 246, groups: Word, 25; Subject, 10

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.63901    1.30857   2.781  0.00542 **
Q2Ayes       0.49931    0.34813   1.434  0.15149
Q2Byes      -1.95048    0.41097  -4.746 2.07e-06 ***
Q3yes        0.93704    0.42152   2.223  0.02622 *
Imitate      0.05317    0.39285   0.135  0.89234
Comment     -3.26204    1.12331  -2.904  0.00368 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Q2Ayes Q2Byes Q3yes  Imitat
Q2Ayes  -0.207
Q2Byes  -0.123 -0.426
Q3yes   -0.442  0.137 -0.269
Imitate -0.518  0.104  0.031  0.559
Comment -0.946  0.043  0.162  0.273  0.290
```

INSIDERS

```
> dat=read.table("Task_B_Insiders.txt", T)
> dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1116 1141   -553     1106
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.32129  0.56683
 Word    (Intercept) 0.18613  0.43143
Number of obs: 1009, groups: Subject, 41; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08402    0.31458  -0.267   0.7894
Awareness    0.26813    0.06838   3.921 8.81e-05 ***
Exposure     0.25507    0.10308   2.475   0.0133 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
        (Intr) Awrnss
Awareness -0.593
Exposure  -0.702  0.078


> dat=read.table("Task_B_Insiders.txt", T)
> dat.lmer = lmer(Answer ~  Q2A + Q2B + Q3 + Imitate + Comment + Travel +
(1|Word) + (1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Comment + Travel + (1 |
Word) +      (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1120 1165 -551.1     1102
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.27500  0.52440
 Word    (Intercept) 0.18553  0.43073
Number of obs: 1009, groups: Subject, 41; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.2109     0.2700   0.781  0.43470
Q2Ayes        0.6347     0.2619   2.424  0.01537 *
Q2Byes       -0.1049     0.2900  -0.362  0.71746
Q3yes         0.5041     0.2845   1.772  0.07638 .
Imitate       0.8237     0.2698   3.053  0.00226 **
Comment      -0.4984     0.2716  -1.835  0.06648 .
Travel        0.6940     0.2514   2.760  0.00578 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Q2Ayes Q2Byes Q3yes  Imitat Commnt
Q2Ayes  -0.404
Q2Byes  -0.266 -0.298
Q3yes   -0.008 -0.104 -0.500
Imitate -0.233  0.034 -0.059  0.040
Comment -0.179 -0.128  0.137 -0.043 -0.519
Travel  -0.396  0.086  0.123 -0.057  0.258 -0.292


dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
```

```
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1116 1141   -553     1106
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.32129  0.56683
 Word    (Intercept) 0.18613  0.43143
Number of obs: 1009, groups: Subject, 41; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08402    0.31458  -0.267   0.7894
Awareness    0.26813    0.06838   3.921 8.81e-05 ***
Exposure     0.25507    0.10308   2.475   0.0133 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
          (Intr) Awrnss
Awareness -0.593
Exposure  -0.702  0.078
```

BILINGUALS

```
> dat=read.table("Task_B_bilingual.txt", T)
> dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
   Data: dat
   AIC   BIC logLik deviance
 337.6 355.7 -163.8    327.6
Random effects:
 Groups  Name        Variance Std.Dev.
 Word    (Intercept) 0.084894 0.29137
 Subject (Intercept) 0.559365 0.74791
Number of obs: 272, groups: Word, 25; Subject, 11

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.3285     0.5596  -0.587    0.557
Awareness     0.1170     0.2331   0.502    0.616
```

```
Exposure      0.3753     0.2327   1.613    0.107

Correlation of Fixed Effects:
        (Intr) Awrnss
Awareness -0.263
Exposure  -0.659 -0.401

dat=read.table("Task_B_Bilingual.txt", T)
> dat.lmer = lmer(Answer ~  Q2A + Q2B + Imitate + Comment +  (1|Word) +
(1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Imitate + Comment + (1 | Word) + (1 |
Subject)
   Data: dat
   AIC BIC logLik deviance
 337.7 363 -161.9    323.7
Random effects:
 Groups  Name        Variance Std.Dev.
 Word    (Intercept) 0.080304 0.28338
 Subject (Intercept) 0.300103 0.54782
Number of obs: 272, groups: Word, 25; Subject, 11

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.1373     0.3052   0.450   0.6527
Q2Ayes        1.3078     0.6043   2.164   0.0304 *
Q2Byes        0.1148     0.7436   0.154   0.8773
Imitate      -1.1359     1.1382  -0.998   0.3183
Comment       1.7815     1.1538   1.544   0.1226
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Q2Ayes Q2Byes Imitat
Q2Ayes  -0.293
Q2Byes  -0.396  0.120
Imitate  0.155 -0.531 -0.611
Comment -0.319  0.296  0.671 -0.866
>
```

MONOLINGUALS

```
> dat=read.table("Task_B_monolingual.txt", T)
```

```
dat.lmer = lmer(Answer ~  Awareness + Exposure + (1|Word) + (1|Subject),
family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Awareness + Exposure + (1 | Word) + (1 | Subject)
   Data: dat
  AIC  BIC logLik deviance
 1348 1374   -669     1338
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.60947  0.78068
 Word    (Intercept) 0.14085  0.37529
Number of obs: 1205, groups: Subject, 49; Word, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.43418    0.36083   1.203   0.2289
Awareness    0.21331    0.08352   2.554   0.0106 *
Exposure     0.03467    0.09693   0.358   0.7206
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
          (Intr) Awrnss
Awareness -0.625
Exposure  -0.645 -0.019



dat=read.table("Task_B_Monolingual.txt", T)
> dat.lmer = lmer(Answer ~  Q2A + Q2B + Q3 + Imitate + Comment +  (1|Word)
+ (1|Subject), family="binomial", data=dat)
> dat.lmer
Generalized linear mixed model fit by the Laplace approximation
Formula: Answer ~ Q2A + Q2B + Q3 + Imitate + Comment + (1 | Word) + (1 |
Subject)
   Data: dat
  AIC  BIC logLik deviance
 1345 1386 -664.6     1329
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.46617  0.68276
 Word    (Intercept) 0.14159  0.37628
Number of obs: 1205, groups: Subject, 49; Word, 25

Fixed effects:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.6133     0.2855   2.148  0.03167 *
Q2Ayes        0.5442     0.2926   1.860  0.06286 .
Q2Byes       -0.4536     0.3158  -1.436  0.15097
Q3yes         0.8207     0.2952   2.780  0.00543 **
Imitate       0.6463     0.2670   2.421  0.01549 *
Comment      -0.5157     0.2747  -1.877  0.06053 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Q2Ayes Q2Byes Q3yes  Imitat
Q2Ayes  -0.445
Q2Byes  -0.119 -0.396
Q3yes   -0.156  0.010 -0.495
Imitate -0.183 -0.008 -0.124  0.135
Comment -0.406 -0.100  0.213 -0.073 -0.413
>
```

WORKS CITED

Aboud, F., & Amato, M. (2001). Developmental and socialization influences on intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell Handbook of Social Psychology: Intergroup Processes* (pp. 65-85). Oxford, UK: Blackwell Publishers.

Anderson, E. (1990). *Speaking with style: The sociolinguistic skills of children*. New York: Basil Blackwell.

Bates, D., Meaechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes, R package version.

Beckman, M., Munson, B., & Edwards, J. (2007). Vocabulary growth and developmental expansion of types of phonological knowledge. In J. Cole & J. Hualde (Eds.), *Laboratoy Phonology 9* (pp. 241-264). New York: Mouton de Gruyter.

Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy, 16*(4), 392-417.

Campbell-Kibbler, K. (2007). Accent, (ING), and the social logic of listener's perceptions. *American Speech, 82*, 34-64.

Chambers, J. (1992). Dialect Acquisition. *Language, 68*(4), 673-705.

Chambers, J. (2002). Dynamics of dialect convergence. *Journal of Sociolinguistics, 6*(1), 117-130.

Clopper, C. (2010). Classification of regional language varieties. In D. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics* (pp. 203-221). New York: De Gruyter Mouton.

Clopper, C., Conrey, B., & Pisoni, D. B. (2005). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology, 24*(2), 182-206.

Clopper, C., Levi, S., & Pisoni, D. B. (2006). Perceptual similarity of regional dialects in English. *Journal of the Acoustical Society of America, 119*, 566-574.

Clopper, C., & Pisoni, D. B. (2004a). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change, 16*, 31-48.

Clopper, C., & Pisoni, D. B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32*, 111-140.

Clopper, C., & Pisoni, D. B. (2007). Free classification of regonal dialects of American English. *Journal of Phonetics, 35*, 421-438.

Cremona, C., & Bates, E. (1977). The developmental attitudes towards dialect in Italian children. *Journal of Psycholinguistic Research, 6*(3), 223-232.

Dailey-O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser like and quotative like. *Journal of Sociolinguistics, 4*(1), 60-80.

Day, R. (1980). The development of linguistic attitudes and preferences. *TESOL Quarterly, 14*(1), 27-37.

Drager, K. (2010). Sensitivity to grammatical and sociophonetic variability in perception. *Laboratory Phonology, 1*, 93-120.

Egerova, N. (2010). *Neural correlates of diaect perception in 5-month-old infants: A Near Infrared Spectroscopy study.* Master's Degree, University of Groningen, Groningen, The Netherlands.

Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year old British children. *International Journal of Behavioral Development, 33*(4), 366-375.

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology, 1*(1), 5-39.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*, 409-438.

Foulkes, P., Docherty, G., & Watt, D. J. L. (2005). Phonological Variation in Child Directed Speech. *Language, 81*, 177-206.

Foulkes, P., Khattab, G., & Yaeger-Dror, M. (2010). Sound Judgments: Perception of Indexical Features in Children's Speech. In D. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics* (pp. 327-356). New York: De Gruyter Mouton.

Fridland, V. (2010). The cycle of production, ideology and perception. In D. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics*. Berlin: De Gruyter Mouton.

Giles, H. (1970). Evaluatve reactions to accents. *Educational Review, 22*, 211-227.

Giles, H. (1971). Patterns of evaluations in reactions to RP, South Welsh and Somerset accented speech. *British Journal of Social and Developmental Psychology, 10*(280-281).

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland & N. Coupland (Eds.), *Contexts of Accomodation* (pp. 1-68). Cambridge, UK: Cambridge University Press.

Giles, H., Harrison, C., Creber, C., Smith, P., & Freeman, N. (1983). Developmental and contextual aspects of children's language attitudes. *Language and Communication, 3*(2), 141-146.

Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology, 26*, 409-433.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865-892.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23*(3), 351-379.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*, 458-484.

Hay, J., Warren, P., & Drager, K. (2010). Short-term exposure to one dialect affects processing of another. *Language and Speech, 53*(4), 447-471.

Hirschfeld, L., & Gelman, S. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development, 12*, 213-238.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & Mullennix (Eds.), *Talker variability in sppech processing* (pp. 145-166). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485-499.

Kerswill, P., & Williams, A. (2000). Creating a new town koine: Children and language change in Milton Keynes. *Language in Society, 29*(65-115).

Kinzler, K., Corriveau, K., & Harris, P. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*(1), 106-111.

Kinzler, K., Shutts, K., DeJesus, J., & Spelke, E. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27*(4), 623-634.

Kinzler, K., Shutts, K., & Spelke, E. (2012). Language-based social preference among children in South Africa. *Language Learning and Development, 8*(3), 215-232.

Kitamura, C., Panneton, R., Notley, A., & Best, C. (2006). *Aussie-Aussie-Aussie, Oi-Oi-Oi: Infants love an Australian accent.* Paper presented at the Conference of the Journal of the Acoustical Society of America.

Labov, W. (1998). The three English dialects. In M. D. Linn (Ed.), *Handbok of dialects and language variation* (pp. 39-81). San Diego: Academic Press.

Labov, W. (2001). *Principles of Socioliguistic Change. Volume 2: Social Factors*. Malden, MA: Blackwell Publishers.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. Berlin: Mouton de Gruyter.

Lambert, W. (1967). A social psychology of bilingualism. *Journal of Social Issues, 23*, 91-109.

McGowan, K. (2011). *The Role of Socioindexical Expectation in Speech Perception.* PhD, University of Michigan, Ann Arbor, MI.

Millar, S. (2003). Children and Linguistic Normativity. In D. a. C. Britan, J. (Ed.), *Social Dialectology: In honor of Peter Trudgill* (pp. 287-297). Philadelphia: John Benjamins.

Munson, B. (2010). Levels of Phonological Abstraction and Knowledge of Socially Motivated Speech-Sound Variation: A review, a proposal and a commentary on the papers by Clopper, Pierrehumbert, and Tamati; Drager, Foulkes; Mack; and Smith, Hall and Munson. *Laboratory Phonology, 1*(1), 157-177.

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language, 25*, 343-365.

Nazzi, T., Jusczyk, P., & Johnson, E. (2000). Language discrimination by English-learning 5-month-olds: Effect of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62-85.

Niedzielski, N. (2010). Linguistic security, ideology and vowel perception. In D. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics*. Berlin: De Gruyter Mouton.

Odato, C. (2010). *Children's development of knowledge and beliefs about English like(s).* PhD, University of Michigan, Ann Arbor, Michigan.

Payne, A. (1976). *The Aquisition of the Phonological System of a Second Dialect.* University of Pennsylvania, Philadelphia, PA.

Phan, J., & Houston, D. (2008). Infant Dialect Discrimination *Research on Spoken Language Processing*. Bloomington, IN: Indiana University, Speech Research Laboratory.

Pierrehumbert, J. (2000). Exemplar dynamics: Word frequency, lenition and contrast *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, acquisition of phonology. *Language and Speech, 46*(115-154).

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics, 34*, 516-530.

Preston, D. (1986). Five visions of America. *Language in Society, 15*(2), 221-240.

Preston, D. (1993). Folk Dialectology. In D. Preston (Ed.), *American Dialect Research* (pp. 333-378). Philadelphia: John Benjamins.

Preston, D., & Niedzielski, N. (1999). *Folk Linguistics*. Berlin: Mouton de Gruyter.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*(1), 10-30.

Rakerd, B., & Plichta, B. (2003). *More on perceptions of /a/ fronting*. Paper presented at the New Ways of Analyzing Variation 32, Philadelphia.

Roberts, J. (1997a). Acquisition of variable rules: A study of (-t,-d) deletion in preschool children. *Journal of Child Language, 24*, 351-372.

Roberts, J. (1997b). Hitting a moving target: Acquisition of sound change in progress by Philadelphia children. *Language Variation and Change, 9*, 249-266.

Roberts, J., & Labov, W. (1995). Learning to Talk Philadelphian: Acquisition of short a by Philadelphia children. *Language Variation and Change, 7*, 101-112.

Rubin, D. (1992). Nonlanguage factors affecting undergraduates judgments of non-native English speaking teaching assistants. *Research in Higher Education, 33*, 511-531.

Schneider, E. (2008). *Varieties of English 2: The Americas and the Carribean*. Berlin: Walter de Gruyter.

Shatz, M., & Gelman, R. (1973) The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development: Vol. 38*.

Silverstein, M. (2003a). Indexical order and the dialectics of sociolinguistic life. *Language and Communication, 23*, 193-229.

Silverstein, M. (2003b). Indexical order and the dialects of sociolinguistic life. *Language and Communication, 23*, 193-229.

Squires, L. (2011). *Sociolinguistic Priming and the Perception of Agreement Variation: Testing predictions of exemplar-theoretic grammar.* PhD, University of Michigan, Ann Arbor.

Staum-Casasanto, L. (2009). What do listeners know about sociolinguistic variation? *University of Pennsylvania Working Papers in Linguistics, 15*(2), 40-49.

Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86-99.

Sumner, M., & Samuel, A. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*, 487-501.

Wells, J. (1982). *Accents of English*. Cambridge, U.K.: Cambridge University Press.