Durian, David. 2004. The social stratification of word-initial (str) in Columbus, OH shopping malls. Manuscript, the Ohio State University. © 2004 David Durian. All Rights Reserved.

The social stratification of word-initial (str) in Columbus, OH shopping malls[1]

David Durian
The Ohio State University

1. Introduction

The following paper reports the results of a pilot study conducted within three shopping malls located in Columbus, OH in an attempt to ascertain whether speakers within the Columbus speech community utilize variant pronunciations of the word-initial consonant cluster (str) and determine if (str) is a salient phonological marker of speech within the community. To study the use of (str), I conducted a rapid anonymous survey of store clerks who work in 29 stores located within these malls, specifically studying the clerks' variable use of two phonetic realizations of (str)--[str] and [Str]--in two speech environments (less emphatic and more empathic speech). As my results will demonstrate, it would appear as though (str) is a salient phonological variable for speakers in the Columbus speech community. As my analysis will also demonstrate, the variable distribution of (str) in the speech of these clerks follows a pattern of usage among the speakers, and its variable occurrence is stratified across the Columbus community by the following stratification social factors: age, sex, and socioeconomic class.

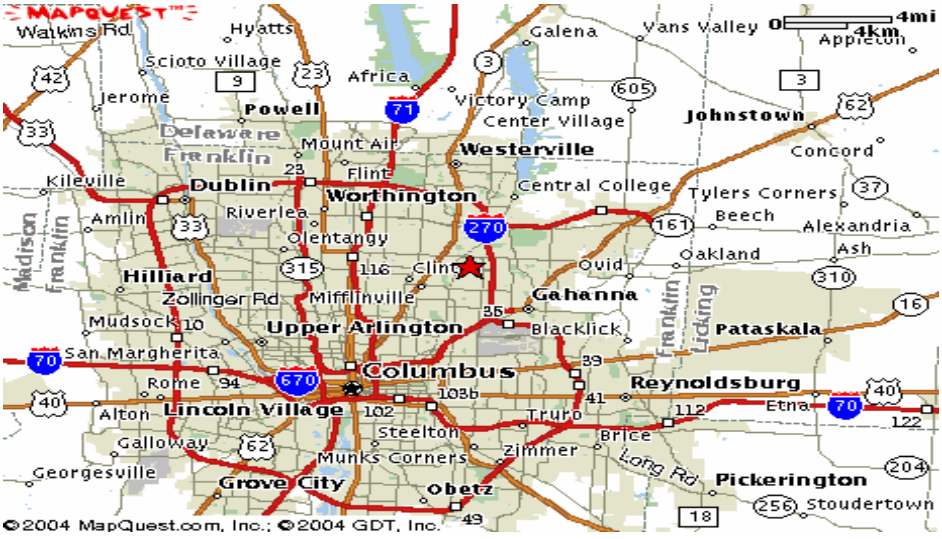
Before discussing these results however, I will first briefly discuss the Columbus speech community, attempting to define the community so that the context within which my results occur can be understood. I will then move to discussing why I have decided to study (str) as the first variable that will feed my large-scale sociolinguistic survey of Columbus, which I plan to undertake during the next 5 years within the community. I then provide an operational definition

[1] I would like to thank Donald Winford for his guidance in the planning and conduct of this study, as well as Robin Dodsworth for her support in helping me code and analyze the class perception survey and GOLDVARB analysis data used in this report. I addition, I would like to thank Stacey Bailey, Na'im Tyson, and Ila Nagar for helping to administer the class perception survey; the students in the three sections of Linguistics 201 at the Ohio State University for completing that survey; and Grant McGuire, Rich Janda, Brian Joseph, and Mike Armstrong for their remarks on the phonetics and phonology of (str) clusters. Any errors which remain are solely my own.

of class as it is used to stratify my data for the purposes of my analysis, the methods I used to obtain the data for my analysis, and the results I obtained via my study, finally moving to the a discussion of these results in the final section.

1.1. The sociolinguistic setting of the shopping mall survey study

Located in the heart of the Midlands dialect region, Columbus is the 15th largest city in the United States and the largest city in the state of Ohio, with a total population of nearly 1.6 million inhabitants (U.S. Census Bureau, 2000). For our present purposes, I am defining the Columbus community as residents living within the bounds of metropolitan Columbus, as well as the surrounding areas currently designated with the boundaries of the Columbus city limits, including unincorporated Columbus. The following study draws on data elicited from store clerks working in the Northern, Eastern, and Metropolitan areas of Columbus, via data collection which occurred at three shopping mall locations in the great Columbus area: Polaris mall, Easton mall, and City Center Mall. The geographic distribution of these general mall locations is highlighted in map 1, below:



Map 1: The geographic distribution of the shopping malls surveyed in Columbus, OH

As Durian (2004) notes, the Columbus speech community has not been studied from a phonological or morphosyntactic vantage point in great detail since the 1930's, when dialectologists surveyed lexical and phonological details of the "Columbus dialect" (McDavid, et al., 1980). Within recent years, several attempts have begun to capture phonological and morphosyntactic aspects of the speech of Columbus residents (Pitt, et al, 2003; Johnson, 2004; Dodsworth, 2004; McGuire, 2003; Weldon, 1993; Armstrong, 2003; Labov, Ash, and Boberg, forthcoming), but as was discussed in Durian (2004), these efforts have been on a relatively small scale across the relevant data sets collected. Because efforts to begin documenting the present day phonological and morphological attributes of the Columbus community have begun so recently, and both diachronic and synchronic coverage of the community has been minor, the need to obtain data on salient phonological variables that are sociolinguistic markers of the speech of Columbus residents of paramount concern so that the large scale study that I have proposed can begin. Hence my choice to study (str) in this pilot survey.

1.2. The choice of the phonological variable (str)

In terms of a choice of phonological variable for study, I have chosen the variable (str) because an exploratory pilot study investigating phonological variance in the speech of Columbus speakers recorded for the Variation in Conversation (ViC) corpus noted that (str) appears to exhibit variable realization as [str] and [Str]among Caucasian middle class speakers living in Columbus. As this study (Armstrong, 2003) noted, the variable (str) occurs widely in the speech of speakers who were recorded for the VIC corpus, a sample of 40 speakers (20 males and 20 females) stratified by sex and age collected several years ago by researchers at the Ohio State university. According to Armstrong (2003), (str) is realized variably as [str] and [Str] in several environments for these speakers—word medially, word initially, and word finally. As

Armstrong, as well as others who have looked at the phonetic environments in which variation in the pronunciation of (str) occurs (Shapiro, 1995; Labov, 2001; Janda and Joseph, 2003) also note, it would appear that speakers utilize variant pronunciations of [atr]and [Str] because the word initial /s/ contained in the cluster undergoes a process of "long distance" assimilation with the /r/ that follows the /t/ in these clusters. This leads to a process where /s/ becomes palatalized, and thus, some speakers variably realize [s] as [S].[2] As my data will demonstrate, and as Armstrong's pilot study as shows, it would appear as though the palatalization process noted in the literature occurs most frequently for speakers who palatalize /s/ in either more rapid or more casual speech environments, while it appears to occur quite infrequently in slower or more emphatic speech environments.

Although Armstong's analysis notes that (str) occurs variably in the speech of the VIC speakers, his analysis stops short of fully accounting for the distribution of (str) among Columbus speakers, because it only takes sex and age into account, and therefore, it does not attempt to correlate the occurrence of the variable with social factors such the socioeconomic status, residential location (neighborhood membership), or ethnicity of the speakers. This is so because the VIC sample was not designed for this type of sociolinguistic research, and therefore, we are unable to determine (str)'s full status as a sociolinguistic marker from Armstrong's analysis. Because this is this case, I have decided to focus on the variable via this large-scale pilot study, to see if a more coherent picture of pattered variable occurrence of (str) can be obtained, based on data which attempts to capture a more complete picture of the social factors (such as the interaction of socioeconomic status, age, and sex of the speakers) [3] which interact in

[2] Although see Lawrence (2000) for a possible alternate explanation of the assimilation process occurring here.
[3] For the purposes of this pilot study, I made the decision to control for race. In the expanded version of this study (Durian and Dodsworth, 2004), we intend to elicit data from African-American Vernacular English speakers in the Columbus community, as well as Spanish Vernacular English speakers.

the variation process.

2. The data elicitation method

I will now move to discussing the data elicitation method I used to obtain samples of (str) tokens from the participants interviewed in my study. To obtain the data, I used a rapid anonymous survey technique modeled after the one used by Labov in his department store study of New York City in 1962 (Labov, 1966; 1972). As Labov (1966) notes, the rapid anonymous survey technique provided invaluable as a pilot study technique in the location of linguistic variables for use in his large-scale study of phonological variation in New York City and for testing out his assumptions about the social stratification and salience of the variable as a marker of New York speech. The same technique provided highly useful of the variable (str) in the study of the Philadelphia speech Labov undertook in the late 1970s (Labov, 1984; 2001), where he was able to use it to trace the path of the variation of (str) across a variety of neighborhoods. Interestingly, Labov's use of the technique in the Philadelphia study also prompted him to hypothesize a trend that has given me additional hope that (str) may be a salient variable in Columbus speech: "Philadelphia shows a systematic advance in the use of hushing sibilants in initial [str/ clusters; again, this appears to be general to many areas of the Midland dialect" (2001:123).

Because of the noted success that researchers such as Labov have had in obtaining data quickly, reliably, and with the use of a minimum amount of recording equipment (all I needed to use was a pencil and a sheet of paper), I decided to use the technique here to obtain my data for this pilot study. Via the use of this technique, I was able to interview 285 Caucasian Standard Vernacular English speaking informants over the course of 4 weekends at four shopping mall areas throughout the Columbus area. Of these 285 informants, exactly 127 provided me with (str)

tokens[4], and of the tokens obtained from this group, I have chosen to analyze the tokens spoken by 120 of these speakers. In terms of total data collection time, I spent roughly 5 hours a day, three weekend days a week over the four weeks collecting my data, for a total data collection time of roughly 60 hours.

From each of the three areas surveyed, I elicited data from 40 informants, using a cell selection method to ensure that a roughly equal number of informants were selected from across the various areas based on the class status (based on the initial 3 level class scale described above) of the stores in which they worked, the sex of the informants, and the perceived age of the informants. The original cell selection method made use of the 3-level socioeconomic class categorization schema referenced above, so that 40 informants (20 male/20 female) where selected to fill the cells for each geographic area, and equal numbers of informants stratified by age were selected across the three areas to yield a total of 120 informants divided into three age classes: 15-30, 35-50, and 55-70. For the purposes of this stage, an estimate of the age of the informant was made, utilizing a five-year window of time to try and obtain a reliable estimate. Data was elicited from informants in two speech environments: the first pronunciation of a (str) variant, which occurred in a less emphatic speech environment, and the second pronunciation of a (str) variant, which occurred in a more emphatic speech environment, since the second pronunciation always occurred when the participant was asked to repeat what he or she had said. Following this approach, 240 (str) tokens (120 less emphatic, 120 more emphatic) were obtained from the participants.

To elicit data from informants, I used one of two elicitation phrases so that comparable

[4] The additional 7 speakers' tokens (14 in total) were not analyzed in this study because they were obtained from speakers at a fourth mall location—Olgentangy Plaza. I ultimately chose not to include this data in my report here due to the limited coverage of the shopping area I was able to obtain within my four-week survey time limit.

data would be elicited from all participants, depending on the shopping scenario the store in question was located within. The elicitation phrase scenarios involved were the following:

Scenario I: Store/Mall Location Near a Major Road Actually Called a Street (for example, Gramcery Street)

In this scenario, the following elicitation phrase was used: "Can you give me directions to ___ avenue?" where the avenue named is actually a "street". When asked for the directions, speakers corrected me by saying "do you mean ___ **street?**" I would then respond "excuse me?" and they would clarify the name of the street by repeating it. Then, respondents would give me directions to the street involved, sometimes providing additional (str) tokens in the process. This method provided me with two (str) tokens in the process—one in a less emphatic environment (first interaction), and one more emphatic (second interaction).

Scenario II: Store/Mall Location Not Near a Major Road Actually Called a Street

Because Polaris Mall was not located within an area containing a major road that is actually called a street, my scenario I question was not able to work within that mall environment. To obtain (str) tokens within the Polaris Mall area, I therefore had to use the following elicitation phrase instead: "Can you tell me where X is?" In this case, X indicates an item in the store or a section of the store (such as the electronics department, the men's shoes section, the home improvement section, etc.) that was located directly behind or across from, the location of the informant in question. When successful, this question elicited the response "Go **straight** across/back that way" from informants. I would then ask the informant to repeat what they said, and therefore, I was able to elicit tokens in comparable environments to the two discussed in scenario I, above.[5]

Data were recorded and coded for store class, speaker age, speaker sex, (str) variant used in both speech environments, occupation of the informant, and regional location of informant based on shopping mall area. As in Labov's (1966) department store study, data were recorded (written down) immediately after the interaction with the participant was completed, once I was out of the line of view of the informant. The data were latter tabulated electronically using Microsoft Excel, so that quantitative analysis of the coded data could be performed.

[5] It should be noted that this method was not as successful as that used in Scenario I, because many participants would either point to the area in question and say something to the effect "go over there," or they would say something like "it's next to the gardening section." Because this method had to be used extensively in the Polaris Mall section, a number of "dropped triggers," in which no (str) tokens were obtained, occurred. This fact explains the relatively low response rate (120 participants out of 285 informants interviewed) obtained in my study.

3. The socioeconomic stratification of the sample

As was eluded to earlier in this paper, being a rapid anonymous survey study, my survey of Columbus shopping mall store clerks takes as its model Labov's 1962 New York City Department store study. Because this is so, the present study makes a number of the same assumptions that Labov's original study did—namely, that macro-level patterns of linguistic variation can be quantifiably studied via the stratificational analysis of the correlation of the observed linguistic variation of a variable (in this case, variant realizations of (str)) with social factors such as the socioeconomic status, sex, and age of the informants interviewed via the survey. Because Labov's model of the analysis of linguistic variation has shown consistent and reliable results in the noting of macro-level patterns of variation, it is my opinion that the adoption of Labov's techniques in this pilot study are justified.

However, as critics of the Labovian stratificational model have pointed out (see, for example, Woodard, 1985; Cameron, 1997; Rickford, 1986), the operationalization of class within Labov's framework is highly problematic, essentially because the categories assigned to stratify the linguistic behavior of informants have often been viewed as one dimensional, flat representations of the social reality in which the participants actually live their daily lives, in which class ranking scales represent an abstract classificatory system which is not salient for the informants studied in a stratificational analysis. Furthermore, as Bell (1984; 2001) points out, Labov's model suffers from a different problem, which is also exhibited by the model of class he used, which is that it essentially misses the interactional nature of the daily lives of the sales clerks, who co-construct their identity with the customers they serve through these sales transactions. In terms of the department store study, although Labov does note that sales clerks are "image conscious" in these transactions in his original work (1972: 46-49), his focus on the

role of class as it is manifested only in the one-sided context of the speech behavior of the store clerks, and not the two-sided context of the impact of the customers on their speech behavior, means that his study failed to take into account this added measure of identity which ultimately affects these clerks' class membership assignments.

In terms of the present study, I will attempt to deal with both of these "class issues" by attempting to build a class raking scale that makes use of Labov's notion of "prestige" into account, since his operationalization of class can help to build an appropriate model to adopt within the specific context of the shopping mall, while also making use of information on the perceived class backgrounds of shoppers who shop at these stores. This is so because, as the store clerks I interviewed indicated to me themselves via informal interviews, they felt that their language use in sales transactions and relations with customers is actually influenced by a prestige-focused model of interaction, which reflected both their feelings (evaluation) about the image of the stores they work for and their evaluation of the customers they serve as store clerks.

My own interviews with store clerks at five stores located throughout the three malls surveyed (Lazarus/Macy's, Kauffmann's, JC Penny's, Wal-Mart, and Nordstrom's) indicates that it is true that the employees of these stores feel that the speech they use in transactions with customers reflects their own feelings on who they feel that the customers who shop in their stores typically are, and they mentioned that they talk differently to customers in the context of sales transactions that might otherwise (say, in the home lives) due to the prestige-centered climate in which they work on a daily basis. This resonates with Labov's discussion of the sales clerks he observed (Labov, 1972: 46-49). During several of the interviews, the sales clerks specifically indicated that they felt the prestige of the store in which they worked factored into this speech behavior during sales transactions, with the store clerks mentioning that they felt that they were

the agents who sold the "image of the store" for which they worked to the customers via the transactions.

Taking this information into account, the ultimate class ranking scale I have utilized for this study draws on three sources of data so that a more "three dimensional" picture of class can be utilized in this study: objective ranking criteria that focuses on the "prestige aspect" of the store clerks speech and allows the stores to be stratified via the criteria—I will refer to this ranking as analyst assigned class; the results of my store clerk interviewers, which lead me to adopting the "prestige model" in this study; and the external evaluative ranking (by 34 college undergraduates) of the perceived class of the "average shopper" who shops at each store, obtained by me through a survey distributed to college students at the Ohio State University, located in Columbus, OH. I will now discuss in some detail how this three dimensional picture of class has been constructed, first focusing on the ranking of the data by the traditional analysts means, and then discussing how the perceived class evaluation component of my ranking system has been constructed. In this way, my study is similar to Labov's in that I ultimately assign class to the stores, with the store clerks working in each store viewed as the enunciators of that class status, as reflected in the speech they use in daily sales transactions.

3.1. The traditional assignment of analyst's categories for class

Following a similar protocol to that used in Labov's department store study (Labov, 1966; 1972), the stores and shopping malls investigated in this analysis were assigned analyst's categories for class based on a number of objective criteria. These criteria were chosen as a way of stratifying the data based on criteria that could be applied across cluster groupings of stores sharing similar class attributes so that cross-group comparisons could be made among the cluster groups.

The criteria used to stratify the data based on socioeconomic class factors are the following:

a) Available advertising and pricing policies as stated on the official Web sites for each store
b) The average price of goods sold in each store (as indicated by price range)
c) The classification of merchandise sold at reduced prices as "clearance" vs. "redline"
d) The spacing of the aisles in department stores and/or the amount of floorspace made available for the display of merchandise sold in each store
e) Average wages paid to employees

An example of how these criteria were used to cluster the stores into class stratified groups is illustrated in table 1, below:

Merchandise Item[6]	Example Stores	Average Price Range
Men's sport shirts	**Sak's Fifth Avenue**	$65.00-$125.00
	Nordstrom's	$65.00-$98.50
	Banana Republic	$49.00-$78.00
	Lazarus/Macy's	$19.00-$90.00
	Old Navy	$12.00-$30.00
	Target	$12.00-$30.00
	Sears	$19.50-$40.00
	JC Penny's	$12.00-$25.00
	Kmart	$9.00-$25.00
Men's casual dress pants	**Sak's Fifth Avenue**	$59.00-$125.00
	Nordstrom's	$65.00-$98.50
	Banana Republic	$40.00-$78.00
	Lazarus/Macy's	$19.00-$90.00
	Old Navy	$19.00-$35.00
	Target	$17.00-$40.00
	Sears	$19.00-$60.00
	JC Penny's	$19.00-$40.00
	Kmart	$12.00-$30.00

Table 1: Example of the selectional criteria used to assign analysts categories to the stores surveyed

As the price ranges indicated in table 1 demonstrate, the average price of goods sold in each store provides with a useful estimator for ranking the stores on a class scale stratificationally. Based on the data contained in this table, we can see that clusters of stores can be formed that allow the stores to be grouped into three groups: stores selling merchandise with high average prices (Sak's Fifth Avenue/Nordstrom's/Banana Republic), middle level average prices (Lazarus/Macy's/Old Navy/Target/Sears), and low average prices (JC Penny's/Kmart).

[6] Prices referenced here were taken from the Web sites of the stores listed during the first week of May, 2004.

Furthermore, as my discussion will now make clear, we will also see that these price point clusters can best be described as being representative of three class ranking clusters: the working class, the lower middle class, and the upper working class.

If we take into account the other factors mentioned in my criteria list above, the distinctions among stores by class that were just mentioned becomes more reified, and their ability to become clustered together in class-based groups is clarified. For example, based on the in table 1, it is hard to tell exactly where one should place JC Penny's in relation to other stores "close by" to it based on merely the price points listed above. However, when we take into account that the amount of floor space dedicated to the display of merchandise is considerably more at Penny's than at stores such as Target, Sears, or Lazarus, and if we take into account that Penny's typically offers more merchandise at clearance prices than these other stores, it becomes clear that Penny's is better categorized as a store clustered among stores in the group of stores more closely aligned with a working class ranking than a lower middle class ranking, as is the case with Target, Sears, or Lazarus. Hence, criteria c, d, and e above serve to clarify the initial distinctions provided by criteria a and b in the assignment of class to the stores surveyed.

Based on the correlation of my informal interview data with store clerks that I mentioned earlier with the objective ranking criteria I have listed above, I feel that the class categories I have assigned to the stores I surveyed via my study is a fairly reliable indicator for stratifying the data based on class attributes. Based on the application of these five selectional criteria listed above, and its correlation with the interview data I have just mentioned, I have divided the 29 stores surveyed in this study into 3 class categories for the purposes of my analysis, with the assigned class categories selected being Working Class, Lower Middle Class, and Upper Middle Class. The division of the stores by class is indicated in table 2, below:

Assigned Class Cluster Title	**Stores Included in Class Cluster**
Working Class (WC)	*Dollar Store, City Center Food Court, the Dollar Store, United Dairy Farmers, Polaris Mall Food Court, Kmart, Wal-Mart, Meijer's, JC Penny's*
Lower Middle Class (LMC)	*The Gap, the Buckle, American Eagle Outfitters, Pacific Sunwear, Gadzook's, Target, Old Navy, Best Buy, Lowes, Lazarus/Macy's, Sears*
Upper Middle Class (UMC)	*Nordstrom's, Kauffmann's, Lord and Taylor, Sak's Fifth Avenue, Anne Taylor, Banana Republic, Eddie Bower, Express (for Men), Express (for Women)*

Table 2: The distribution of stores surveyed, clustered by analyst assigned class categories

Ultimately, based on this type of analysis, the stores are able to be ranked by objective criteria and broken in to clusters based on class. As a result, the store clerks, and representatives of these stores, are viewed as being representative speakers of the social classes depicted. Thus, on this account, the store clerks within the sample population become stratified in the same way as Labov's store clerks were (1966; 1972). As I will now show, this model of class provides half of the picture for the complete class model used in this study. The second half, which attempts to account for the other half of the interactional picture I described early, will now be discussed.

3.2. The ranking of perceived class by external evaluators

As was mentioned early, to gage the class background of the customers who typically shop at the stores which I surveyed, I decided to poll 34 undergraduate college students, so that the perceived class of these customers could be determined. Since the store customers shopping in the stores at the time of my survey themselves could not be polled in an objective manner directly (without me revealing my presence as a researcher), and since the store clerks I surveyed might be biased in their descriptions of the customers who shop in these stores, I decided to use

these undergraduate students' evaluations, since they themselves are an active "target population" served by these Columbus stores. In this way, a more complete picture of the class status of the stores, based on the interactional model I discussed early, could be obtained.

The perceived class survey administered to the classes consisted of a seven level class ranking scale (Lower Working Class, Middle Working Class, Upper Working Class, Lower Middle Class, Middle Middle Class, Upper Middle Class, Upper Class), and the students were asked to rank a sample of 35 stores based on this criteria.[7] Seven levels of class category were used for the survey in an attempt to reduce potential bias on the raters. I decided to do this because I wanted the raters to feel "they had options," which is something I felt would not have been possible if they were restricted only to the three scale ranking system I used in my analysts category analysis.

To administer the class survey, the help of three course instructors of three undergraduate linguistics courses at the Ohio State University was enlisted.[8] The surveys were distributed by the course instructor to his or her students, the purpose of the survey was explained, and students were given the chance to "opt out" of the survey if they wished. Once students agreed to take the survey, the survey was completed by the students, and the results were calculated by me using Microsoft Excel. The survey is attached as Appendix A of this report.

Once the results of the survey were tabulated, the stores covered in the survey were clustered into "class blocks" based on the modal distribution of the tallies obtained. A model distribution method was used because it allowed the stores to be ranked based on frequency clustering, with the highest amount of category tally points used to establish the central class

[7] Not every store I obtained data from was included in the survey, and additional stores were also listed. This was so because I wanted to obtain a general picture of the students' perceptions of customer class in these stores, as a way of attempting to account for evaluator bias.

[8] The results of only two of the classes' survey rankings is reported here. Results from the third class were omitted due to time constraints that did not permit me to code their results.

rank position for the cluster analysis. Thus, if a store received 34 rankings, and 20 tally points were placed in the Middle Middle Class column, while only 10 points were placed in the Lower Middle and only 4 points were placed in the Upper Middle Class column, the store would be assigned to the Middle Middle Class cluster based on this distribution.

Based on this survey method, the following ranking results for the stores actually surveyed in the rapid anonymous survey, were obtained[9]:

Assigned Class Cluster Title	**Stores Included in Class Cluster**
Lower Working Class (LWC) – Middle Working Class (MWC)	*Dollar General , the Dollar Store*
Middle Working Class (MWC)-Upper Working Class (UWC)	*United Dairy Farmers, Meijer's, Wal-Mart, K-Mart*
Upper-Working Class (UWC)-Lower Middle Class (LMC)	*Gadzook's, Sears, JC Penny's, Target*
Lower-Middle Class (LMC) –Middle Middle Class (MMC)	*Express for (Men), Express (for Women), The Gap, American Eagle Outfitters, Pacific Sunwear, Lowe's, Old Navy, Kauffmann's, , the Buckle, Lazarus*
Upper Middle Class (UMC)	*Nordstrom's, Lord and Taylor, Sak's Fifth Avenue, Eddie Bower*

Table 3: The distribution of stores surveyed, clustered by perceived class categories

As the data in table 3 illustrate, a more diverse class ranking emerged when participants were given the choice to rank the stores themselves. As a result of the modal analysis of the participants' responses, five categorical distinctions arose: LWC-MWC, MWC-UWC, UWC-LMC, LMC-MMC, and UMC. Based on initial analysis conducted via the multivariate analysis

[9] The raw statistical data for these results are given in Appendix B of this paper.

program GOLDVARB, it was determined that the five-level system presented in table 3 above was more significantly correlated with the other factors analyzed in this study (speech environment, sex, age, and department store region) than the 3 level class ranking schema used in the analyst-assigned section of this paper.

Because this was so, when the time came to merge the two class analysis approaches together into a unified class scale, so that a more "three dimensional model" of class could be constructed, I made the decision to cluster the stores based on the 5-level structure presented in table 3. Following the clustering of the stores together, the following cluster groups emerged:

Assigned Class Cluster Title	**Stores Included in Class Cluster**
Lower Working Class (LWC) – Middle Working Class (MWC)	*Dollar General , the Dollar Store, City Center Food Court, Polaris Mall Food Court*
Middle Working Class (MWC)-Upper Working Class (UWC)	*United Dairy Farmers, Meijer's, Wal-Mart, K-Mart, JC Penny's*
Upper-Working Class (UWC)-Lower Middle Class (LMC)	*Gadzook's, Sears, Target*
Lower-Middle Class (LMC) –Middle Middle Class (MMC)	*Express for (Men), Express (for Women), The Gap, American Eagle Outfitters, Pacific Sunwear, Lowe's, Old Navy, Kauffmann's, the Buckle, Lazarus/Macy's*
Upper Middle Class (UMC)	*Nordstrom's, Kauffmann's, Lord and Taylor, Sak's Fifth Avenue, Anne Taylor, Banana Republic, Eddie Bower*

Table 4: The distribution of stores surveyed, clustered by both perceived class and analyst-assigned categories

In terms of the results reported in section 4 of this paper, the five-level class ranking schema presented above will be used to classify the stores in which the store clerks worked and ultimately assign class to the store clerks themselves. As this merged class structure above

reveals, a class-ranking system that accounts for a more three dimensional, interactional model of social class has been obtained through the merger of my analyst-assigned rankings and the student evaluator perceived class category rankings discussed in this section. Now that we have an understanding of how class has been operationalized in this study, I shall now move to discussing the results of my rapid anonymous survey.

4. The results

In the following section, I will discuss the interaction between (str) and the individual social factors (independent variables) chosen for analysis in this study, namely: socioeconomic class (combined scale), sex of the informant, perceived age of the informant, general mall location, and speech environment (less emphatic vs. more emphatic). As we will see, almost all of these social factors play an important role in the stratification of (str) in Columbus shopping malls, although as the multivariate analysis at the end of this section will reveal, not every one is, in the end, statistically significant.

I have decided to employ both a simple statistical analysis comparing the ratio of [str] vs. [Str] and a multivariate regression analysis of the co-variance of realizations of (str) as it interacts with the social factors described above. As this analysis provides a more detailed account of the distribution of (str) within the sample population, the results of this analysis ultimately are the most important that I report here.

In the charts used throughout this paper, the following coding system is used to describe the variable occurrence of [str] vs. [Str] in my data:

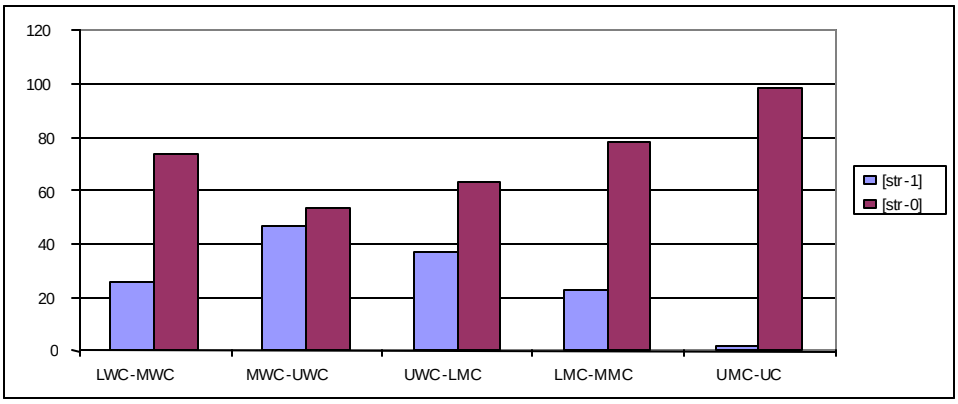
[str-0]: the informant used the variant [str]
[str-1]: the informant used the variant [Str]

To begin my discussion of the results, I will first briefly discuss the simple statistical analysis of individual social factors

4.1. Distribution of (str) by social class

As figure (1) illustrates, the occurrence of (str) across both speech environments investigated in this study (less emphatic vs. more empathic) shows a stratificational pattern when the social class of the speakers surveyed is taken into account. As the figure illustrates, employees of the middle to upper working class stores surveyed used [ʃtr] variant the most frequently, using this variable 47% of the time (18 out of 38 tokens) when the total number of (str) tokens elicted in my survey is taken into account. The upper working class-lower middle class closely follows middle to upper working clerks in usage, with these clerks using [ʃtr] 37% of the time (11 tokens out of 30). In terms of frequency of occurrence, the lower to middle working class (LWC-MWC) employees and the lower middle to middle-middle class (LMC-MMC) employees use the next greatest frequency of [ʃtr] across speech environments, with the LWC-MWC clerks using [ʃtr] 26% of the time (9 out of 34 tokens) and the LMC-MMC clerks using it 23% of the time (18 out of 80 tokens). Finally, the upper middle class/upper class speakers trail far behind these other classes, using [ʃtr] only 2% of the time (only 1 token out of 58).

As the percentage usage scores below indicate, the occurrence of [ʃtr] is highly stratified, with the pattern of variation exhibited following closely the "classic pattern" of variation one might expect in a pilot survey such as this: [ʃtr] patterns exactly along the lines of Labov's curvelinear hypothesis (2001), with the MWC-UWC and UWC-LMC using [ʃtr] far most frequently among the groups observed. Thus, it would appear that these patterns indicate that the lower the class of the speaker, the more likely he or she is to use the [ʃtr] variant, while the higher the class of the speaker the less likely he or she is to use [str]. As the results of the multivariate analysis in section 4.6 will reveal, the social class effect is statistically significant.

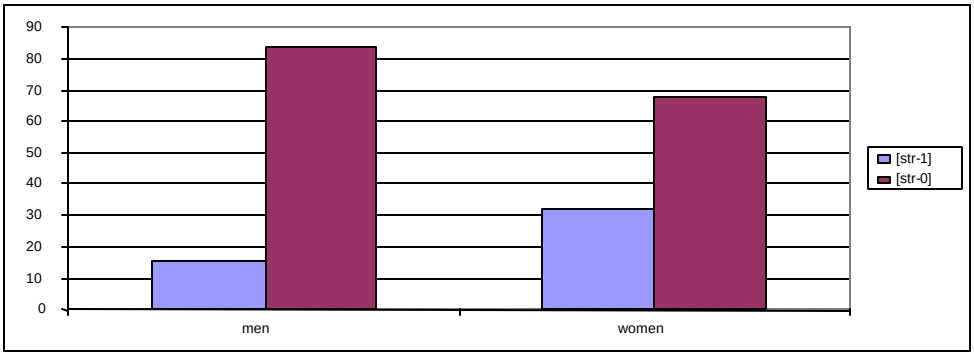


	26%	**74%**	**47%**	**53%**	**37%**	**63%**	**23%**	**78%**	**2%**	**98%**
# of tokens	**[9/34]**	**[25/34]**	**[18/38]**	**[20/38]**	**[11/30]**	**[19/30]**	**[18/80]**	**[62/80]**	**[1/58]**	**[57/58]**

Figure 1: Distribution of (str) by combined class scale [n= 240 tokens total]

4.2. Distribution of (str) by sex

As figure (2), below, illustrates, the occurrence of (str) across both speech environments also shows a differentiated pattern when the sex of the store clerks is taken into account. The women surveyed used the [ʃtr] the most frequently, using this variable 32% of the time (38 out of 120 tokens).On the other hand, men used [ʃtr] only 16% of the time (19 out of 120 tokens).



	16%	**84%**	**32%**	**68%**
# of tokens	**[19/120]**	**[101/120]**	**[38/120]**	**[82/120]**

Figure 2: Distribution of (str) by sex [n= 240 tokens]

This means that women used [ʃtr] roughly 2 times more than men in my survey population. As my discussion of the findings in section 4.6 will reveal, this is another statistically significant pattern of variation.

4.3. Distribution of (str) by perceived age of informant

In figure (3), below, we see that the occurrence of (str) across both speech environments also shows stratification by perceived age. As figure (3) reveals, when age is taken into account, the youngest speakers surveyed used [ʃtr] the most frequently, using this variable 34% of the time (27 out of 80 tokens). As this figure also reveals, participants in the 35-50 age range used [ʃtr] 29% of the time (23 out of 80 tokens), and so their usage of [ʃtr] closely resembles the usage of speakers in the 15-30 year old range. Finally, as figure (3) also reveals, the oldest informants surveyed only used [ʃtr] 9% of the time (7 out of 80 tokens), which indicates there is a radical decrease in the usage of [ʃtr] for participants older than 55 in the sample population. This quite surprising figure shows us that speakers under the age of 55 used [ʃtr] nearly 3 times as much as participants over the age of 55, a figure which may indicate that [ʃtr] is a recent innovation in the speech of the Columbus population. Again, section 4.6 will reveal that this pattern is statistically significant.

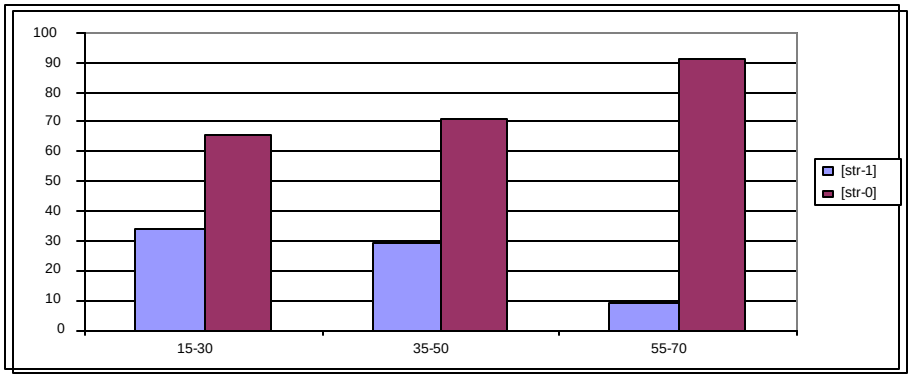
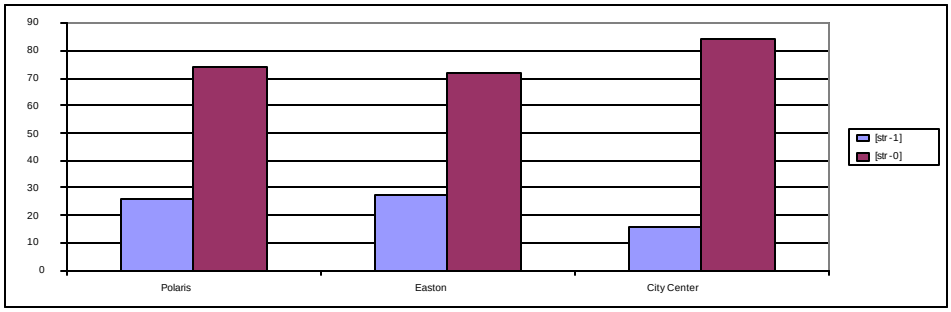


Figure 3: Distribution of (str) by perceived age of informant [n= 240 tokens]

4.4. Distribution of (str) by shopping mall area

As figure (4), below, reveals, the occurrence of (str) across both speech environments is only weakly conditioned by geographic area. Store clerks working in the Polaris Mall area used

[ʃtr] the most frequently (38% of the time (23 out of 80 tokens)), with clerks working in the Easton Mall area following closely behind their usage (26% of the time, or 21 out of 80 tokens). In contrast, informants working in the City Center mall area used [ʃtr] 16% of the time (14 out of 80 tokens), and so their usage of [ʃtr] does not appear to differ significantly from the usage patterns exhibited by clerks within the other shopping mall areas. What these figures ultimately seem to suggest is that the region in which the shopping mall is located does not have a pronounced effect on the frequency of usage of [ʃtr] in the speech of my store clerk informants. This finding is verified further via the results of the multivariate analysis of social factors, which I will discuss in section 4.6 below.

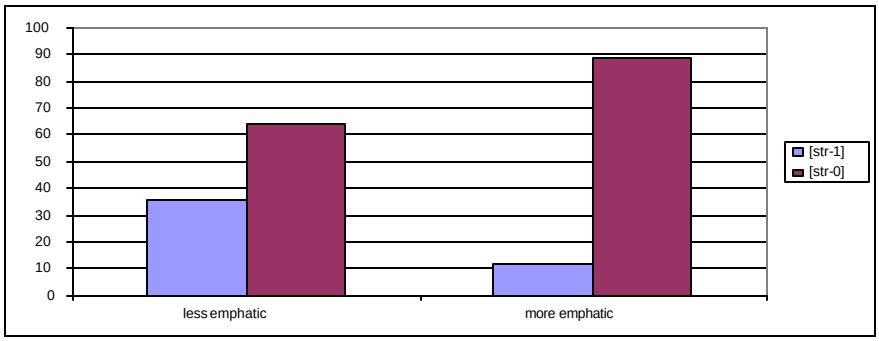


	Polaris	Easton	City Center
	28% 72%	26% 74%	16% 84%
# of tokens	[23/80][57/80]	[21/80] [59/80]	[14/80] [66/80]

Figure 4: Distribution of (str) by shopping mall area [n= 240 tokens]

4.5. Distribution of (str) by speech environment

Perhaps not surprisingly, when the data is categorized by the two speech environments surveyed in my data set, we find that there is a marked increase in the usage of [ʃtr] in comparison to [str], with [ʃtr] occurring 3 times as frequently for all participants surveyed in the study in the less emphatic environment than in the more emphatic environment. This fact is revealed by figure (5) above, which shows us that speakers used [ʃtr] in less emphatic speech environments 36% of the time (43 out of 120 tokens), while they used [ʃtr] only 12% of the time (14 out of 120 tokens) in the more emphatic speech environment.



	less emphatic		more emphatic	
	36%	**64%**	**12%**	**88%**
# of tokens	**[43/120]**	**[77/120]**	**[14/120]**	**[106/120]**

Figure 5: Distribution of (str) by speech environment [n= 240 tokens]

As will now be discussed, this finding is also revealed to be statistically significant via the multivariant regression analysis of variation.

4.6. The multivariant analysis of the variable occurrence of (str)

The results of the GoldVarb analysis support the generalizations I have discussed above. The following independent variables are shown by this analysis to be statistically significant (p=.001): class, sex, age, and speech environment. As these results also show, and was as previously discussed, geographic region turns out to not be significant. The overall results of a binomial 1-level analysis (after recoding to exclude or regroup statistically insignificant factors) are shown in table (5), below.

Concerning this data, two facts should be noted about the clustering of the factor groups shown in table (5). First, it should be noted that speakers within the two “under 55” age groups were clustered together, as the results of earlier GoldVarb runs indicated that this factor group would be best analyzed by making the grouping distinction noted in the table (in other words, clustering the 15-30 year speakers and the 35-50 year old speakers into one group). Second, it should be noted that speakers working within the Polaris Mall and Easton Mall regions were also clustered together, again due to the their similar weights in an earlier run.

Log Likelihood -86.816				
Significance .001				
Input 0.102				
Chi-square/Cell .7579				
Group	Factor	Weight	App/Total	Input & Weight
1: Gender	M	0.334	0.016	0.05
	W	0.666	0.032	0.18
2: Environment	Less Emphatic	0.731	0.36	0.24
	More Emphatic	0.269	0.12	0.04
3: Age	15-55*	0.663	0.31	0.18
	55-70	0.206	0.09	0.03
4: Region	Polaris/Easton*	0.562	0.27	0.13
	City Center	0.376	0.16	0.06
5: Social Class	MWC-UWC	0.863	0.47	0.42
	LWC-MWC	0.711	0.26	0.22
	UWC-LMC	0.740	0.37	0.24
	LMC-MMC	0.584	0.22	0.14
	UMC-UC	0.061	0.02	0.01

Table 5: Results of the GOLDVARB Analysis

Overall, the results of the GOLDVARB analysis overlap heavily with the results of the simple statistical analysis of individual factors discussed in sections 4.1-4.5 above. However, what the GOLDVARB analysis lets us see more clearly is the overall statistical significance of the co-variance of these factors when they are considered together as a group. Specifically, they help us to more confidently predict the environments in which the clerks' use of [ʃtr] will occur. Based on the interaction of the input value (0.102) and the factor weights revealed for each factor, we see that women are predicted to use [ʃtr] 18% of the time, speech environment not withstanding. As well, we see that all speakers are most likely to use [ʃtr] in the less emphatic speech environment, with GOLDVARB predicting that these speakers will use [ʃtr] 24% of the

time within this environment. In addition, we see that younger speakers are predicated to use [ʃtr]18% of the time in both speech environments, and that UWC-LMC speakers are predicated to use [ʃtr] 42% of the time in both environments. Because the Chi-Square/Cell value is .7579, we can be fairly confident that these predications are valid for the speech sample population studied in this survey.

5. Conclusion: Some possible explanations for the distribution of (str)

Based on the results reported in section 4 of this paper, there appear to be two possible explanations that explain the distribution of (str) variants in this sample. The first is that patterns of variation noticed reflect differences in the pronunciation of (str) in more casual speech of the informants versus their use of (str) variants in more formal speech—in other words, these patterns reflect that (str) is not a variable engaged in a "change in progress," but instead, (str) is a stable sociolinguistic variable. This type of explanation suggests that [str] would appear to be the more carefully pronounced variant of (str), based on the extremely high occurrence of [str] in the more emphatic context, and a lower (but still pronounced) usage in the less emphatic context. Based on this analysis, for example, it would appear that the younger WC or LMC female speakers observed in this study were merely using variant pronunciations of (str) in the two speech contexts because these contexts so readily focus on their variant pronunciations. Because these speakers would be less concerned with using a more careful pronunciation in the first context observed, they would be more prone to using the less careful pronunciation of (str) in this environment. However, in the more emphatic environment, where the speaker thought that I had not understood them and would therefore need to be told directions or a street name more carefully, they would naturally use the more carefully pronounced variant [str], so that the words they were pronouncing were made more explicitly clear.

Evidence supporting the "careful" vs. "more casual" analysis of the data can be drawn from the results reported above themselves—namely, the pronounced shift in the use of [str] in the more emphatic speech environment, and the more pronounced use of [str] among all speakers of all classes save for the MWC-UWC and the UWC-LMC blocks. On this analysis, the social factors found to correlate with these variant pronunciations, although significant, would be downplayed in the overall explanation of the patterning of the variance exhibited by the data, since the speech environments in which "more casual" vs. "more careful" speech would most heavily account for the variance observed. A second source of evidence supporting this view is that we currently do not have enough evidence, drawn from either the Columbus speech community nor other speech communities, to yet determine if another possible explanation (say, that the variable distribution of (str) is a sound change in progress), based on the lack of historical evidence concerning the "introduction" of a variants of (str) that might be involved in a change in progress. The strongest evidence for this argument comes from Lawrence (2000), who argues that the variant pronunciation of (str) might be traced as far back as the 1850s, and therefore, by this point in history (2004), any changes in progress that might be associated with (str) would have long since been completed.

A second possible explanation of the variable occurrence of (str) within my data set would be that (str) is in fact a change in progress, and therefore, the relatively low occurrence of the [ʃtr] pronunciation of (str) might demonstrate that this variable is a recent innovation for speakers in the Columbus speech community, and that it's introduction into the greater speech community is still be negotiated. On this view, [ʃtr] would be treated as the new variant, while [str] would be treated as the older variant. Speakers such as younger women, specifically women working within the MWC-UWC and the UWC-LMC stores, would be viewed as the possible

innovative change agents, and they would be the speakers of most interest in investigating further to track the development of this change in progress.

Evidence to support this hypothesis comes again from the data observed via this study. Specially, the radical change in the occurrence of usage among speakers who are younger than 55 seems to indicate that (str) might be a recent innovation in the community, perhaps introduced within the last 30 to 40 years, since female speakers within the 35-50 age group show significant variance in the use of variant pronunciations of (str). Additional evidence that this is a change in progress, and one that might be spreading, is presented by the increase among 15-30 year old women in the usage of [ʃtr] when their usage is compared with the 35-50 year old female speakers. As well, although their production of [ʃtr] is markedly less noted, younger male speakers within the survey population also appear to engaging in the use of the newer form, as the evidence presented specifically in sections 4.2, 4.3, and 4.6 would suggest. Because the possible pattern of innovation suggested by these types of results fits the "classical" theoretical model of a sound change in progress typically presented in the sociolinguistic literature (see, for example, Labov (1972) and (2001) and Janda and Joseph (2003)) I cannot rule out that this analysis of the occurrence of (str) in Columbus shopping malls is valid.

Of course, this view does have its problems, based on the patterns also revealed by my data. One notable problem is that speakers shift so heavily their usage of [str] to [str] in the more emphatic context observed in this paper. In sum, only 14 speakers overall actually used [ʃtr] in both speech contexts. If the variant realization of [ʃtr] is a change in progress, and [ʃtr] is in fact the newer variant, we might expect to see more usage of [ʃtr] across both speech contexts for the younger speakers. Therefore, we cannot totally rule out the possibility, based on the data here,

that variable realizations of (str) by these speakers is NOT evidence indicating a change in progress.

However, it should be noted that, regardless of which possible explanation of the distribution of (str) in this survey data, we can be fairly confident in saying, based on the data presented in this paper, that the variant use of [str] shows interesting distributional patterns based on the social class, sex, and age of the speakers observed which merit further study if we wish to make sense of the variable use of (str) in the Columbus speech community. The overall conclusion that can be drawn at this point from the data presented in this paper is that the variable use of (str) is salient within the Columbus speech community, and therefore, it does appear to be a phonological variable of some importance in "Columbus speech" as well as possibly the Midlands dialect speech area. The findings of this study, taken in conjunction with Labov's comments on his study of the distribution (str) in Philadelphia (Labov (1984) and (2001)) provide the tentative evidence to suggest this possibility for Midlands dialect speech.

If we are ultimately to make sense of the puzzle presented by the variance in (str) usage presented in this paper, it would appear that we need to obtain more data on its usage across a number of additional speech contexts, from a more diverse group of subjects (including speakers from various racial backgrounds), and from a larger group of speakers overall, in an attempt to hone in on the patterns of variance which have been tentatively noted via this analysis. This is a matter I will take up in future research (Durian and Dodsworth (2004) and Durian (forthcoming)) which I plan to undertake this summer in the community. In doing so, I hope to be able to more confidently assess the status of (str) in the Columbus speech community, and validate one or the other of the possible hypotheses put forth in these concluding remarks—either that (str) is a variable undergoing a change in progress or that (str) is a stable sociolinguistic variable.

References

Armstrong, Michael. 2003. /s/-retraction in the ViC corpus. Paper presented in Linguistics 802, *Historical Morphology*, the Ohio State University.

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

Bell, Allan. 2001. Back in style: Reworking audience design. In Eckert, Penelope, and John Rickford. (Eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, pp. 139-169.

Cameron, Deborah. 1997. Demythologizing sociolinguistics. In Coupland, Nikolas, and Adam Jaworski (Eds.), *Sociolinguistics : A reader*. New York: St. Martin's Press. pp. 55-69.

Dodsworth, Robin. 2004. Attribute networking: A sociolinguistic technique for modeling social perceptions. Draft in progress, the Ohio State University.

Durian, David. 2004. A preliminary overview of the language change and variation in Columbus, OH (LCVCO) project. Unpublished manuscript, the Ohio State University.

Durian, David. forthcoming. Some remarks on linguistic variation in Columbus, OH: Results of a pilot study investigating phonological variation. Draft in progress, the Ohio State University.

Durian, David, and Robin Dodsworth. 2004 (forthcoming). Urbanization and the spread of linguistic change: (str) in Columbus. Paper prepared for NWAV(E) 33, Ann Arbor, Michigan.

Janda, Richard and Brian Joseph. (2003). Reconsidering the canons of sound change: Towards a "big bang" theory. In *Historical Linguistics 2001. Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne, 13-17 August 2001*, ed. by Barry Blake and Kate Burridge. Amsterdam: John Benjamins Publishing Co. pp. 205-219.

Johnson, Keith. (forthcoming). Massive reduction in conversational American English.

Labov, William. 1966. *The social stratification of English in New York City.* Arlington, VA: Center for Applied Linguistics.

Labov, William. 1972. The social stratification of (r) in New York City department stores. In *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press. pp. ?? -??.

Labov, William. 1984. Field methods of the Project on Language Change and Variation. In Baugh, John, and Joel Schezer (eds.), *Language in use*. Englewood Cliffs, NJ: Prentice Hall. Pp. 28-53.

Labov, William. 2001. *Principles of linguistic change, volume 2: Social factors*. Oxford: Blackwell.

Labov, William, Sharon Ash, and Charles Boberg. forthcoming. *The Atlas of North American English.* Mouton/de Gruyter.

Lawrence, Wayne P. 2000. /str/ --> /Str/: Assimilation at a distance? *American Speech*, 75.1: 82-87.

McDavid, Raven I., Raymond K. O'Cain, George T. Dorrill, and Guy S. Lowman, Jr. (eds). 1980.*Linguistic Atlas of the Middle and Southern Atlantic States*. Chicago: University of Chicago Press.

McGuire, Grant. 2003. The realization of interdental fricatives in Columbus, OH, AAVE. Paper presented at the Montreal-Ottawa-Toronto Phonology Workshop, Toronto.

Pitt, Mark, Keith Johnson, Elizabeth Hume, Scott Kiesling, and William Raymond. forthcoming. The ViC corpus of conversational speech. Manuscript submitted to IEEE Transactions on Speech and Audio Processing: Special Issue on Spontaneous Speech Processing.

Rickford, John. 1986. The need for new approaches to social class analysis in sociolinguistics. *Language and Communication, 6*.3: 215-221.

Shapiro, Michael. 1995. A case of distant assimilation: /str/ —> /Str/. *American Speech, 70*.1: 101-107.

U.S. Census Bureau. 2000. *United States Census 2000*. Available online: http://www.census.gov

Weldon, Tracey. 1994. Variability in negation in African-American Vernacular English. *Language Variation and Change, (6)*:3, 359-397.

Woodard, Kathern. 1985. Language variation and cultural hegemony: Toward an integration of linguistic and sociolinguistic theory. *American Ethnologist 12: 738-748*.