

Language
Variety
in the South
Revisited

Edited by

Cynthia Bernstein

Thomas Nunnally

Robin Sabino

The University of Alabama Press
Tuscaloosa and London

The Role of Social Processes in Language Variation and Change

Jan Tillery

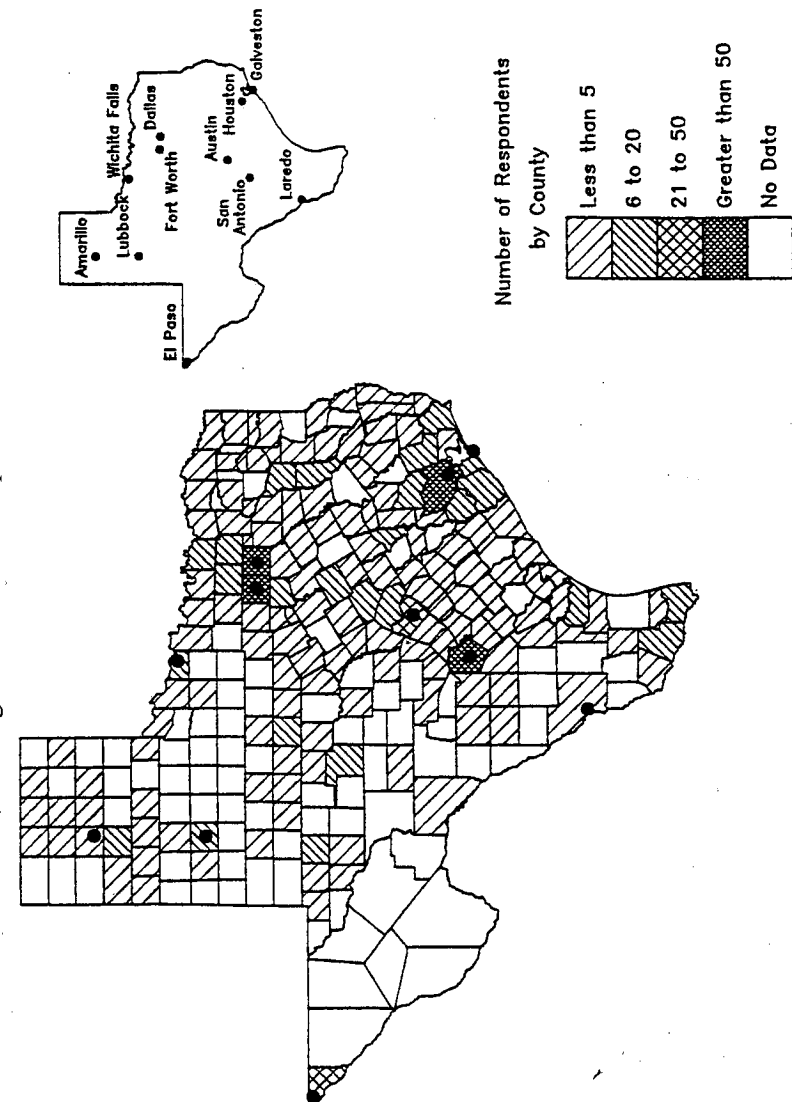
INTRODUCTION

Traditionally, linguists have examined linguistic variation from two perspectives, that of dialect geography and that of sociolinguistics. Each field has developed distinct approaches for gathering and analyzing data and ultimately for explaining variation. On the one hand, dialect geographers, following Kurath and his predecessors in Europe, are concerned with the areal distribution of linguistic features and the dialect areas delineated by those distributions. In order to determine these distributions, dialect geographers typically lay out a grid of communities based on the settlement history of an area, select native informants whose roots in that community often go back three generations, and interview those informants with a detailed questionnaire which (in the United States anyway) links one atlas project with another. From these interviews, dialect geographers plot the spatial distribution of linguistic features and link that distribution to settlement history (Pederson 1972). Even allowing for changes in technology, recent dialect geography surveys are quite similar to those begun a half century earlier.

Sociolinguists, on the other hand, examine the distribution of linguistic variants among social groups and link variation to cleavages in social structure. In order to locate variation in the social structure, sociolinguists—following Labov (1966, 1972b), Milroy (1987a), and others—conduct either random sample surveys or in-depth community studies. Sociolinguists then correlate respondents' ethnicity, social class, age, and gender with linguistic features to explain the linguistic variation that they find. In many ways the methods of sociolinguists have become as fixed as those of dialect geographers. Social class, gender, and ethnicity have become as conventional as isoglosses. Wolfram's (1991) textbook on American dialects, which offers sections on social class, gender, and ethnicity as well as region as correlates for language variation, clearly shows how these variables have become fixed. While dialect geography and sociolinguistics continue to offer crucial insights into language variation, our research in Texas and Oklahoma suggests that neither approach provides a complete picture.

What the data from Texas and Oklahoma suggest is that factors not usually given prominence in either approach actually are as important in influencing variation as the traditional ones. Bernstein (1993, 1994), in her work on data from the Phonological Survey of Texas (PST), suggests that variation according to region, gender, and ethnicity account for a relatively small portion of the variation of that corpus, while our work in Oklahoma shows the impact of such features as rurality and nativity. In other words, the work in Texas and Oklahoma shows that we need to expand the set of explanatory categories used by dialect geographers and sociolinguists. This paper begins such an expansion by reporting on linguistic variation that occurs in the data from a Survey of Oklahoma Dialects (SOD).

Figure 1: Location of respondents in PST



METHODS

The data used in this research come from three sources: a 1989 Phonological Survey of Texas (PST), a 1989 Grammatical Investigation of Texas Speech (GRITS), and a 1991 Survey of Oklahoma Dialects (SOD).¹ PST, GRITS, and SOD are all large-scale multifaceted investigations which include field and random sample telephone surveys of Texas and Oklahoma. The telephone surveys serve as the central component of each and provide the corpus of data for this paper. Figures 1 and 2 show the location of respondents of the telephone surveys in both states.

PST, GRITS, and SOD have been developed in an attempt to gather valid and reliable synchronic linguistic data to which a wide range of statistical and cartographic procedures can be applied.² For each respondent, these surveys elicit the standard demographic data—age, ethnicity, gender, occupation, levels of income, and education—for correlation with the linguistic variables. Figures 3–6 summarize the data for both states. A significant difference occurs in the ethnic makeup of both states as illustrated in figure 4. The Black and Hispanic populations are much greater in Texas than in Oklahoma, but the Oklahoma sample includes 6.9% Native Americans.

Another striking difference between the two, as figure 7 demonstrates, is the proportion of the population in large metro areas. Nearly half of the Texas respondents reside in metropolitan areas of over one million (Dallas/Ft. Worth, Houston, and San Antonio), while nearly three-quarters of the Oklahoma respondents reside in areas under 100,000. These differences are crucial for explaining the patterns of variation that occur in each state.

In addition, the Texas Poll, of which PST was a part, included a question asking respondents to rate Texas as a place to live. The correlation of responses to this question with the linguistic data suggested to us that affective factors might play an important role in explaining variation. For example, our analysis shows that the correlations between the use of monophthongal /aɪ/ and the standard social categories are confusing at the least; however, the use of monophthongal /aɪ/ and respondents' rating of Texas correlate strongly with one another. As figure 8 indicates, those respondents who rate Texas positively (excellent or good) as a place to live are much more likely to use monophthongal /aɪ/ before voiceless obstruents than those who rate the state as fair or poor. In other words, monophthongal /aɪ/ reflects a strong identity with the state (Bailey 1991; Tillery 1990; Tillery and Bailey forthcoming-a). Variation in the use of monophthongal /aɪ/, then, seems to be associated with a respondent's identity with place, a factor (or variable) not measured by most standard social categories. With this in mind, SOD was designed to explore the identity of respondents by gathering crucial information about respondents' perceptions of their regional, state, and local identities and their own status in relationship to them.

For example, we asked respondents how they would rate Oklahoma as a place to live and whether they view Oklahoma as a Southern, Midwestern, or Western state. Here, most Oklahomans, nearly 85%, rate the state favorably as a place to live (see figure 9), while over half consider the state midwestern, and one-third consider it to be southern as figure 10 shows.

Figure 2: Location of Respondents in SOD

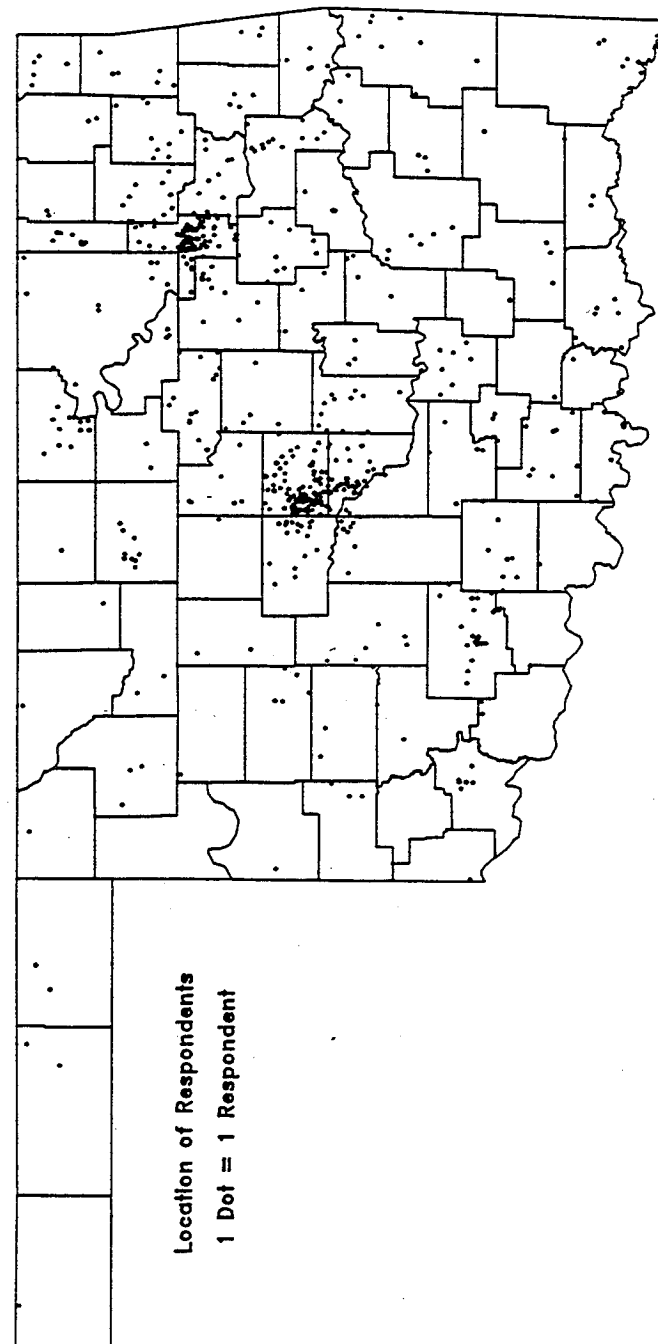


Figure 3: % of total number of PST and SOD respondents by age group

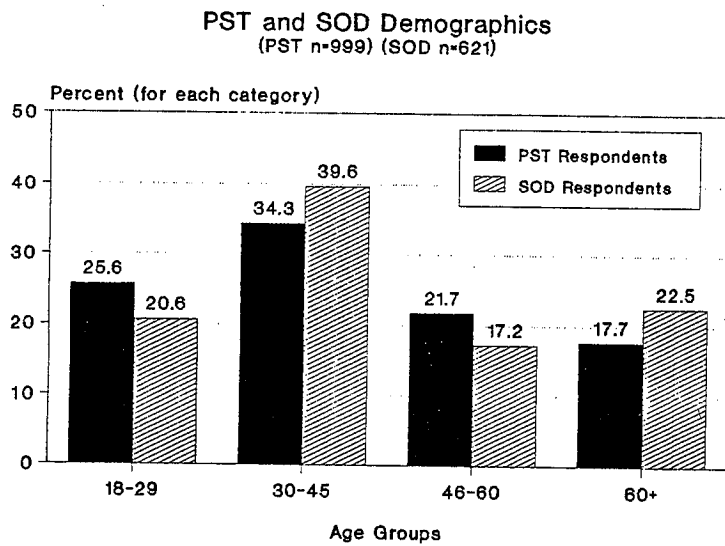


Figure 4: % of total number of PST and SOD respondents by ethnic groups

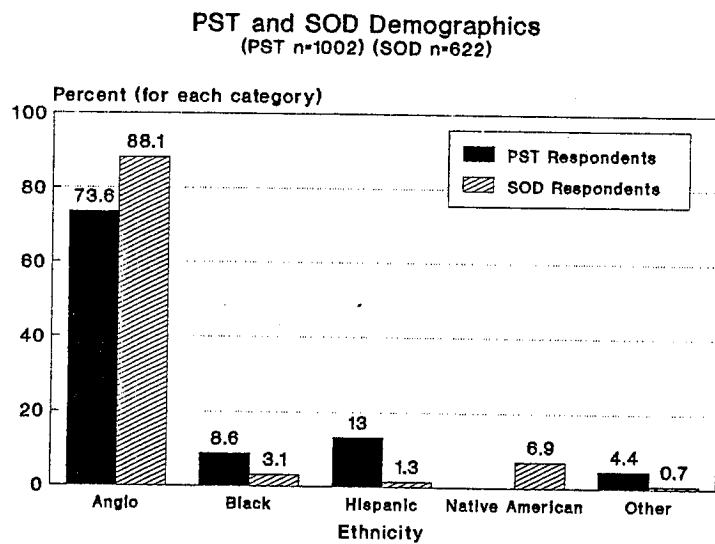


Figure 5: % of total number of PST and SOD respondents by gender

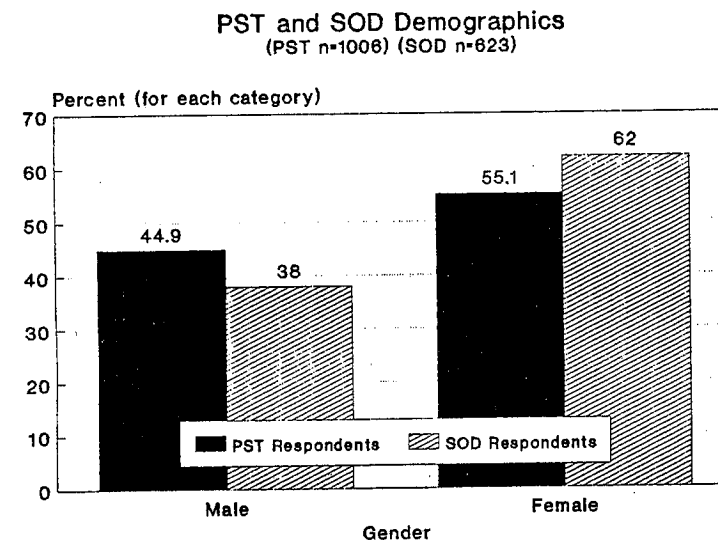


Figure 6: % of total number of PST and SOD respondents by education level

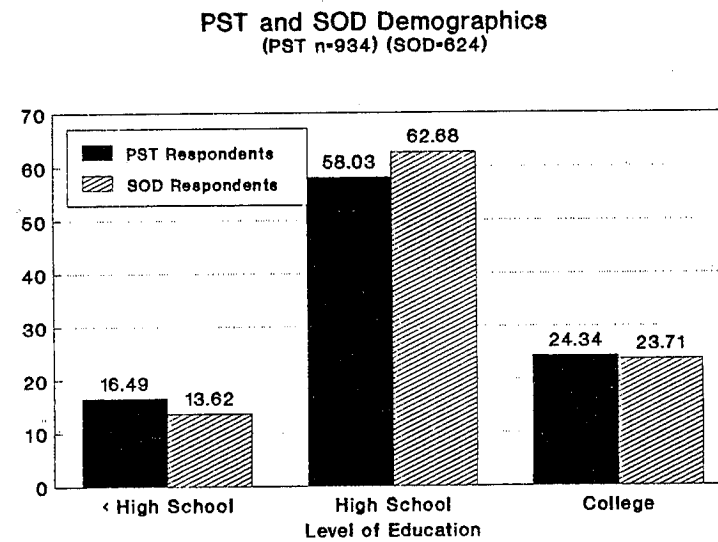


Figure 7: Urban/rural distribution of PST and SOD respondents

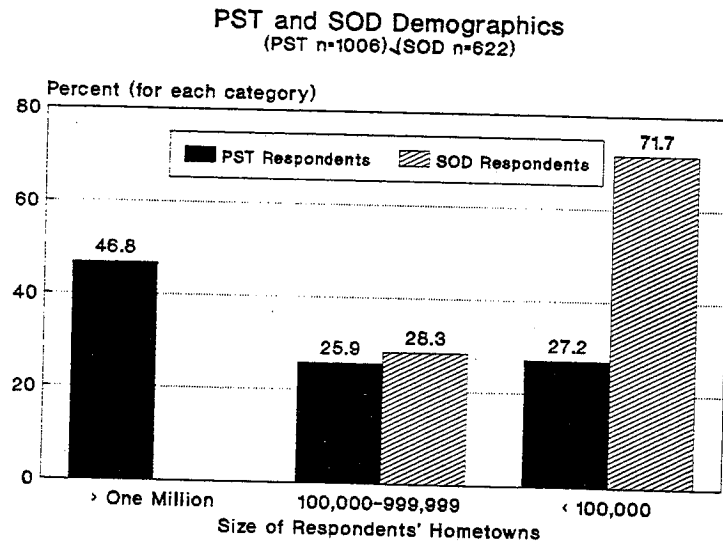


Figure 8: Correlation of monophthongal /aɪ/ in *night* with respondents' rating of Texas

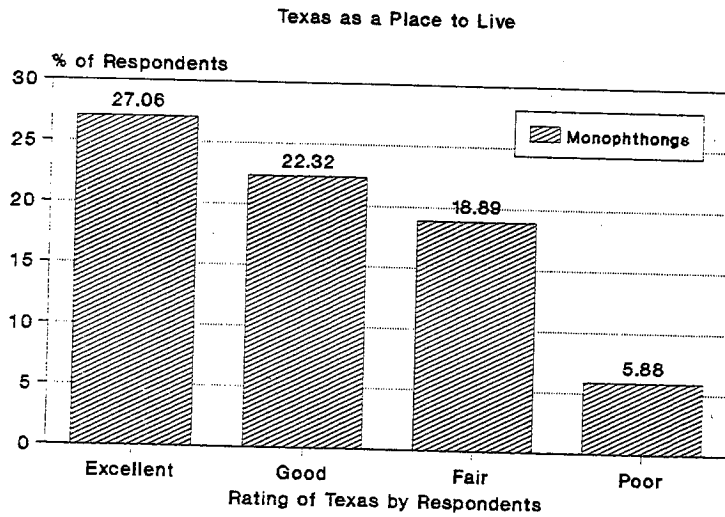


Figure 9: Rating of Oklahoma as a place to live

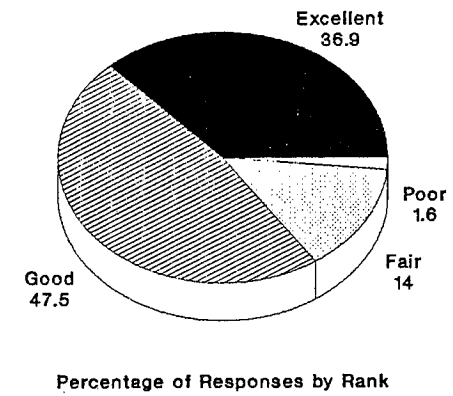
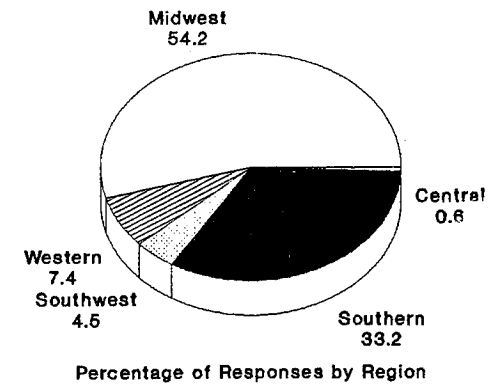
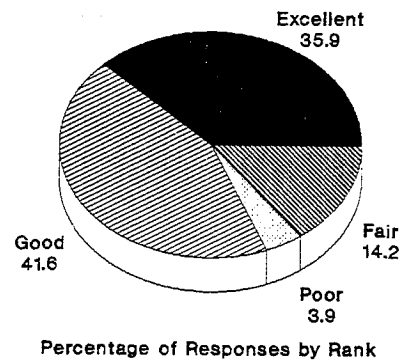


Figure 10: Perception of Oklahoma's location



Following this question, respondents were asked to rank their neighborhoods as places to live. Figure 11 illustrates that Oklahomans rate their neighborhoods almost exactly as they rate the state, a consistency which demonstrates the strength of local identity in Oklahoma.

Figure 11: Neighborhood rating



We explored the effect of neighborhood in one other way as well. Respondents were asked how long they had resided in their current neighborhoods and the size of the communities where they had resided most of their lives. These elaborated measures go beyond nativity (length of residence in the state) and rurality (size of current residence) in that they enable us to assess the effect of the social process of mobility.

RESULTS

A comparison of the effects of standard social categories with the effects of the expanded measures of nativity and rurality demonstrates the importance of mobility on linguistic variation in Oklahoma. The distribution of linguistic variants across the categories of age, gender, occupation, income, education, and ethnicity accounts for only a small portion of the variation in the SOD data.

Table 1 presents all of the phonological variables in SOD, along with an indication of the statistical significance (using the chi square test) of their distribution according to standard sociolinguistic categories. Note that only three of the fourteen linguistic variables are affected significantly by at least three social factors. If age is eliminated, only fifteen of the 84 cells (less than 20%) are filled. Gender, occupation, and ethnicity are significant for only two features each, while income is significant for only three. Education, the category that affects the greatest number of linguistic features, influences only six, less than half of the phonological variables investigated in SOD. While standard social factors are sometimes impor-

tant in explaining variation in Oklahoma, they do not tell the whole story. If we correlated only these six social variables with the phonological variables in SOD, we would have to conclude that much of the phonological variation in SOD is random.

Table 1: Statistical significance of correlation between social categories and SOD phonological variables

Variable	Social Category					
	Age	Gender	Occupation	Income	Education	Ethnicity
/j/ in <i>Tuesday</i>	—	—	—	—	—	—
/ʒ/ in <i>Thursday</i>	—	—	—	—	—	.01
/ə/ in <i>forty</i>	.05	—	—	—	—	.01
/æ/ in <i>thousand</i>	—	—	—	—	—	—
/ə-/ in <i>wash</i>	.01	—	—	.05	—	—
/ɪ/ in <i>Wednesday</i>	—	—	—	—	.01	—
/ɪ/ in <i>pen</i>	—	—	—	—	.01	—
/a/ in <i>Friday</i>	—	—	—	—	—	—
/a/ in <i>time</i>	.01	—	.05	.01	.01	—
/a/ in <i>night</i>	—	—	—	—	.05	—
/ɪ/ in <i>field</i>	.01	—	—	—	—	—
/ɛ/ in <i>bale</i>	.01	.01	—	—	—	—
/u/ in <i>pool</i>	.01	—	—	.05	.01	—
/a/ in <i>hawk</i>	.01	.05	.05	—	.01	—

However, correlating phonological variables with the measures of nativity and rurality reveals that the variation is systematic, not random at all. Table 2 provides a striking contrast to table 1. Of the 56 cells in table 2, 25 (44%) are filled. In other words, nativity and rurality are significant factors over and over again. For three of the phonological features (intrusive [r] in *wash*, a high front vowel in *pen*, and a lax vowel in *pool*), all four of the demographic variables are statistically significant. At least two of the demographic variables are significant for five other features as well (a high front vowel in the first syllable of *Wednesday*, monophthongal /aɪ/ in *Friday*, *time*, and *night*, and an unrounded vowel in *hawk*). These demographic variables are not significant for only three features: the presence or absence of /j/ in *Tuesday* and uncontracted [r] in *Thursday* and *forty*. However, the latter two variables are ethnic features that occur almost exclusively among African-Americans. The presence or absence of /j/ in *Tuesday* represents an anomaly that correlates with none of the social variables investigated in SOD. What the data in tables 1 and 2 clearly show is that social class, ethnicity, and gender play relatively minor roles in variation in Oklahoma. The factors that seem to influence variation most in Oklahoma are how long respondents have lived in the state and in their neighborhoods and whether their communities are urban or rural.

While the data from PST and GRITS suggest that length of residence in the state is a key factor in linguistic variation in Texas, the data from SOD suggest that length of residence in a local neighborhood is more important in Oklahoma. For three of the five features in table 2 (intrusive [r] in *wash*, high front vowel in *pen*, and a lax vowel in *pool*), length of residence in one neighborhood is at least as

significant as length of residence in the state. Moreover, for three variables (lax vowels in *field* and *bale* and an unround vowel in *hawk*), length of residence in the neighborhood is significant, while length of residence in the state is not. These distributions strongly suggest the importance of local identity in language variation in Oklahoma.

Table 2: Statistical significance of correlation between categories of nativity and rurality and SOD phonological variables

Linguistic Variable	Nativity	Years in Neighborhood	Rurality	Size of Longest Residence
/j/ in <i>Tuesday</i>	—	—	—	—
/ɜ/ in <i>Thursday</i>	—	—	—	—
/ə/ in <i>forty</i>	—	—	—	—
/æ/ in <i>thousand</i>	.05	—	—	—
/ə/ in <i>wash</i>	.05	.05	.05	.05
/ɪ/ in <i>Wednesday</i>	.01	—	.01	.01
/ɪ/ in <i>pen</i>	.01	.01	.01	.01
/a/ in <i>Friday</i>	—	—	.01	.01
/a/ in <i>time</i>	—	—	.01	.01
/a/ in <i>night</i>	—	—	.01	.01
/ɪ/ in <i>field</i>	—	.05	—	—
/ɛ/ in <i>bale</i>	—	.05	—	—
/u/ in <i>pool</i>	.05	.01	.01	.01
/a/ in <i>hawk</i>	—	.01	.01	—

Note: Nativity = length of residence in Oklahoma

Rurality = size of place of current residence

Rurality is an equally important factor in language variation in Oklahoma. Rurality is statistically significant (again by chi square tests) for eight of the fourteen variables in table 2. Here, the size of the current community (not the community where respondents have lived the longest) seems to be the key. For example, in every instance that size of the place of longest residence is significant, size of current residence is also statistically significant. In fact, in the use of unround vowels in *hawk* size of current residence is significant while size of the place of longest residence is not. While the effects of nativity and rurality might seem to be unrelated, together they provide evidence on the most important social process of this century—the development of widespread geographic mobility.

Over the last hundred years (and especially the last fifty) the American South has been characterized by extensive movement from the country to the city. This urbanization paralleled migration from the South to the North between World War I and the 1960s and over the last twenty years paralleled migration from north to south as the Sunbelt became an important destination for business and industry, tourism, and retirement. When considered together the four variables of size of current residence, size of the place of longest residence, length of residence in the state, and length of residence in the local neighborhood allow us to measure respondents' participation in the processes of urbanization and migration. The chi square distributions in tables 1 and 2 above suggest that it is these processes of urbaniza-

tion and migration, rather than cleavages in social structures, that motivate language variation and change in Oklahoma. The speech of those who have moved, whether from the country to the city or from state to state, tends to be different from the speech of those who have not.

While urbanization and migration are the chief correlates of language variation in Oklahoma, that is not necessarily the case everywhere. Table 3 illustrates this point. A chi square analysis of the correlation between the phonological variables in PST that appear also in SOD and the social variables of PST shows a much more complex interaction. Of the eight social variables in table 3, all but two significantly affect three or more phonological variables. While gender and income seem to have little influence on variation in Texas, occupation and education are about as influential as nativity and rurality. It is important to keep in mind here, however, that length of residence in Texas is the only measure of nativity in PST, while size of current residence is the only measure of rurality. More elaborate measures, like those done in Oklahoma, might provide different results. The effects of ethnicity are more widespread than any of these: ethnicity significantly affects six phonological variables. As the demographic comparison of Oklahoma and Texas above illustrates, Texas is an ethnically more complex state; linguistic variation in the state reflects that fact. The fact that a substantial number of Texans live in large metropolitan areas suggests that the state may be more complex socially than Oklahoma as well. Certainly the effects of occupation and education suggest that complexity. The important point here, though, is that the relative influence of social variables differs from one state to another and possibly one community to another.

Table 3: Statistical significance of correlation between social and phonological variables in PST

Linguistic Variable	Social Variable							
	Age	Gender	Occ.	Income	Educ.	Ethnic.	Nativity	Rurality
/j/ in <i>Tuesday</i>	.0001	—	—	—	.006	—	.04	.0001
/ə/ in <i>forty</i>	—	.05	—	—	—	.0001	—	—
/æ/ in <i>thousand</i>	—	—	—	—	—	.0001	—	—
/ə/ in <i>wash</i>	.0001	—	.01	—	—	—	—	—
/a/ in <i>night</i>	—	—	—	—	—	.01	—	—
/ɪ/ in <i>field</i>	.001	—	.0001	—	—	—	—	.002
/ɛ/ in <i>sale</i>	.0001	—	.001	.05	—	—	—	.03
/v/ in <i>school</i>	.0001	—	.0001	—	—	.003	.001	.009
/a/ in <i>lost</i>	.0001	—	—	—	.04	.0001	.0001	—
/a/ in <i>walk</i>	.0001	—	—	—	.05	.0001	.0001	—

The work of Bernstein (1993) on data from PST points to an even more fundamental problem with the standard explanatory variables used in dialect geography and sociolinguistics. Bernstein looks not just at the statistical significance of social variables but at their interaction with one another and at the amount of variation that they actually explain. The results are enlightening. For example, a bivariate analysis of the Texas Poll data from PST shows that the five linguistic features in her cluster 1 are significantly affected by age, ethnicity, income, nativity,

region of Texas, and rurality. Bernstein's multivariate analysis shows, however, that when interactions among factors are taken into account, only age and rurality have a significant effect. Moreover, these two factors only explain 25% of the observed variance in the sample.

Two surprising facts become clear from Bernstein's analysis. First, conventionally used social variables such as gender and social class contribute little to variation in the Texas Poll data, although ethnicity and region are powerful effects for some clusters. Age is the only factor which significantly affects all clusters. Second, we must look to variables such as nativity and rurality to fill in the explanatory gap. However, it is not enough simply to include nativity and rurality in our analyses. Rather we must be open to a wide range of social and demographic processes and not limit our analyses to standard categories that emerged from the early benchmark studies.

CONCLUSION

While gender, occupation, income, and ethnicity will remain important parts of the explanatory apparatus, those categories should not become the "be all and end all" of sociolinguistic analysis. By the same token, we should not be limited by the standard apparatus of dialect geography, i.e., standard isoglosses and dialect areas. Our research in Texas and Oklahoma suggests that spatial distribution of linguistic forms is often quantitative and that spatial distribution tends to interact with social factors in intricate ways to create complex linguistic landscapes (see Bailey, Wikle, and Tillery, 1992). To understand the interaction of the social with the spatial, we need multidimensional models that show those interactions.

With all of these social and spatial categories, we need to remember that they are only oblique indicators of social processes, such as migration and urbanization, that underlie those categories. In developing categories for sociolinguistic analyses, we need to seek those categories that are tied to social processes as directly as possible. Perhaps the most important thing our research in Texas and Oklahoma demonstrates, however, is that preconceptions limit discovery. What the data in Texas revealed with regard to the effects of respondents' identities to Texas on language variation was so dramatic that we expected a duplicate of that in Oklahoma. This did not occur. In the Oklahoma data, mobility is the key factor. Whatever motivation may lie behind linguistic variation elsewhere, one thing is certain—only the data know.

NOTES

1. PST, GRITS, and SOD have been funded by grants from the National Science Foundation (BNS-8812552, BNS-9009232, and BNS-9109695). I wish to thank Guy Bailey for his insights and suggestions during this study.

2. For in-depth discussions of the methodologies used in PST, GRITS, and SOD, see Bailey and Bernstein (1989); Bailey and Dyer (1992); Bailey, Wikle, and Sand (1991); Bailey, Wikle, Tillery, and Sand (1992); Bernstein (1994); and Tillery (1992).

An Ethnolinguistic Approach to the Study of Rural Southern AAVE

Patricia Cukor-Avila

ETHNOGRAPHIC RESEARCH

Central to the linguistic study of any community is a clear understanding of the relationship that exists between its inhabitants and the sites of linguistic interaction within that community, basically who talks to whom and where. In order to develop this understanding the linguist must also be an ethnographer, as it is only through participant observation that the linguist can truly observe the interaction between language and social life, or what Hymes (1964a) calls the "ethnography of speaking." Sociolinguistic research over the past two decades has benefited greatly from the combination of the qualitative methods of ethnographic analysis and the quantitative methods of linguistic analysis. Milroy (1987b) in Belfast, Eckert (1989a) in Detroit, Cheshire (1982) in Reading, Bortoni-Ricardo (1985) in Brazil, Edwards (1986) in Great Britain, and Gal (1979) and Lippi-Green (1989) in Austria are just a few who have successfully implemented participant observation in order to obtain what they feel is the most "naturalistic" data from the communities studied. Ethnographic description has also been the basis of more descriptive linguistic analysis, such as Blom and Gumperz's (1986) work on code-switching in Hemnesberget, Norway, and Schieffelin's (1990) study of the Kaluli and their language.

This paper offers a method for gathering large amounts of linguistic data in a context that preserves natural sociolinguistic interaction.¹ It does so by describing my research in Springville, Texas, over the past five years. By applying insights from ethnography to sociolinguistic fieldwork, I demonstrate how sociolinguists can shift the focus of study to the community itself, allowing the recording of people as they normally interact with each other on a daily basis and therefore the study of language as a "socially situated cultural form" (Saville-Troike 1989:3).

SPRINGVILLE

Springville is a small East-Central Texas town of fewer than 200 people.² It is an ideal laboratory for developing field methods because it is a stable, self-contained community whose social relationships are easily definable and whose sites of linguistic interaction are easily identified. Moreover, it provides an ideal laboratory for studying rural, insular varieties of African-American Vernacular English (AAVE) since it preserves almost intact the plantation system of social and economic organization that developed during slavery and persevered through the emergence of the tenant farming system. Although this system has disappeared in most places, it remains largely in place in Springville.