Towards a Sociologically-Grounded View of Occupation in Sociolinguistics

Jon Forrest
Robin Dodsworth
Towards a Sociologically-Grounded View of Occupation in Sociolinguistics

Abstract
In order to improve our operationalization of class in sociolinguistic analysis, this paper draws on sociological theory as the foundation for a new approach to the conception and coding of occupation. The 162-speaker dataset is drawn from the larger corpus of sociolinguistic interviews conducted in Raleigh, NC. F1 and F2 measurements for the five front vowels of the SVS were extracted at 25% vowel duration and Lobanov-normalized (1971), and the vowel diagonal (Z2-Z1) was included as the dependent variable in regression analyses. To operationalize a sociologically-based theory of occupations, we implement a five-way distinction between industrial/occupational sectors (Law and Government, Technology and Finance, Interactive Service Work, Care Work, and Blue Collar) based on historical changes in Raleigh’s economy. Net of social and linguistic controls, models show significant differences between groups formerly grouped together as White Collar occupations, attributable to historical embeddedness in the greater Raleigh area.
Towards a Sociologically-Grounded View of Occupation in Sociolinguistics

Jon Forrest and Robin Dodsworth

1 Introduction

Sociolinguists have long been dissatisfied with existing approaches to social class (Ash 2013), citing the lack of engagement with sociological theory and the associated failure to explain the quantitative relationships between linguistic variables and socioeconomic characteristics (Rickford 1986). Central to many class indices in sociolinguistics is occupation, which Labov (2001) argues to be the most important social determinant of class stratification, greater than either educational attainment or income. To implement occupation in sociolinguistic analyses, Labov’s (1972) early work used a typology drawn from census data (Duncan 1961), categorizing occupations along an ordinal socioeconomic status (SES) scale. However, even with the foregrounding of occupation when dealing with class, relatively little attention has been paid to the process by which occupations gain their association with prestigious or vernacular linguistic features. Recognizing this need, work using the linguistic marketplace framework (Sankoff and Laberge 1978) found strong correlations with three Montreal French variables when categorizing occupations by their need for standard linguistic features. Despite these results, the linguistic marketplace approach has rarely been implemented since (Horvath 1985), and it provides little insight into the process underlying the association between standard language and jobs. Even class typologies drawn from ethnographic data (Eckert 2000), while consistent with the stratification system of the community (Baranowski 2008, Hazen 2011), do not necessarily speak to the mechanisms by which linguistic variables develop correlations with socioeconomic characteristics.

With the goal of building a class paradigm grounded in contemporary sociological research, we draw upon two strains of literature. First, we look to research on the occupational structure of the United States, which has increasingly focused on the social aspects of different types of work (Dwyer 2013, Hochschild 1983, Leidner 1993) over census-style typologies (Duncan 1961, Nakao and Treas 1992). Second, we employ research within economic sociology on the operation of labor market processes (Lin 1999, McDonald and Elder 2006) and the embeddedness of the economic realm within social relations (Granovetter 1985, Polanyi 1957). As a test case for our application of sociological theory, we use the Raleigh Corpus of conversational interviews. Previous work with this dataset (Dodsworth 2014, Forrest 2015) has found a retreat over time from Southern vocalic features, with white collar speakers leading the changes (Dodsworth 2013, Dodsworth and Kohn 2012). This study re-examines the class effects at work in Raleigh, NC, paying special attention to differences between industries that either have historical significance in the area or are main drivers of social inequality.

2 Creating a Sociological Schema for Occupation

2.1 Changing Industries in the U.S. Economy

Rather than utilizing traditional blue collar/white collar divisions like those of foundational studies (Duncan 1961), recent research in the sociology of work narrows its focus to individual industries, exploring how their internal workings contribute to the reproduction of inequality. With the decline of manufacturing jobs and subsequent rise of the service industry (Lorence 1991, Wright and Dwyer 2003), research on service work has identified an increasing ghettoization of low-wage occupations in the field. These jobs, termed “interactive service work,” entail highly routinized, heavily monitored tasks on a day-to-day basis, with monetary exchange at the heart of any interaction (Leidner 1993). Low-level sales work and fast-food workers typify this category, since every customer interaction is so highly scripted, and they receive very low wages for their labor. These jobs have replaced middle-wage factory and manufacturing jobs (Lorence 1991), taking the place of blue-collar work within the US economy and creating an even wider income inequality gap (Wright and Dwyer 2003).

The care industry has also experienced dramatic growth since the mid-20th century, and like
interactive service work, it is mostly comprised of lower-wage work. Though similar to service work in its focus on interpersonal interaction on the job, care work places more emphasis on emotional labor (Hochschild 1983) rather than exchange of material goods. A child-care worker is a prototypical example of a care industry employee, since the job tasks focus on emotional concerns (nurturing, teaching, keeping children happy) rather than the monetary exchange that lies at the heart of other service work. While care work shares the wage penalty associated with service work, it has a much greater degree of gender-typing. Care-oriented jobs are overwhelmingly in female-dominated fields (Duffy, Albelda and Hammonds 2013, Dwyer 2013), and they therefore suffer a further wage penalty regardless of required experience or educational attainment (England 1992, 2005, England, Allison and Wu 2007).

2.2 Structural Embeddedness and Industry Growth in Raleigh, NC

In addition to industry-specific research on occupations, we also employ a theoretical perspective drawn from economic sociology termed “embeddedness,” or the notion that all economic activity is embedded within a broader set of existing social relations (Granovetter 1985, Polanyi 1957). This idea has been implemented in previous sociolinguistic research as a measurement of an individual’s connectedness within local social networks (Dodsworth 2014), but we take a slightly different perspective here. Since our focus is occupation, we are concerned with an industry’s historical embeddedness within the Raleigh, NC area, which reflects the underlying network structures of an industry’s employees. Individual firms tend to hire employees who are culturally similar to existing workers (Rivera 2012), and in general, job leads garnered from network contacts tend to result in a higher likelihood of hire and a better starting wage (McDonald and Elder 2006, McDonald, Lin and Ao 2009). The net result of these processes is that individuals with higher social capital (Portes 2000) (i.e., well-placed network contacts) tend to experience more labor market success over the course of a lifetime (McDonald 2011). The crucial importance of network contacts when getting a job results in the exclusion of certain social groups over others, reinforcing material inequalities (DiMaggio and Garip 2012, Tilly 1999). Ultimately, all these individual-level processes have effects within larger organizations, like firms or industries.

Historical embeddedness matters for this corpus due to the rapid population increase experienced in the greater Raleigh area, which was largely driven by the growth of specific industries. The establishment of Research Triangle Park (RTP) in the 1960s attracted a number of technology companies to Raleigh, bringing with them employees from other branches in the Northern United States. The dialect contact situation resulting from the influx of non-Southern speakers has caused a general decline of traditional Southern features for younger Raleigh residents (Dodsworth 2013, 2014, Dodsworth and Kohn 2012). Importantly for this analysis, the relocation of technology industries from other parts of the country implies a very different network and hiring structure at work in the recruitment and retention of employees. If employers tend to hire culturally similar employees through social networks (McDonald and Elder 2006, Rivera 2012), it would stand to reason that employees would have linguistic similarities as well, being embedded in social networks with non-Raleigh residents. Therefore, the decline of Southern features would likely proceed more quickly for workers in this industry compared to other industries, since less-Southern dialects would more closely match the linguistic features of technology employers. Conversely, Raleigh has a long history of being a law and government town, with prominent jobs in state government and legal fields. Within this industry, we would expect the decline of Southern features to proceed more slowly, since these jobs are so deeply embedded within local networks.

The incorporation of sociological perspectives on national economic changes and structural embeddedness provides an opportunity to look at occupations in more detail within the Raleigh dataset. When implementing traditional occupational coding schema, white collar speakers lead in the changes away from the Southern Vowel Shift. Upon closer examination of the data, however, the white collar group exhibits considerable interspeaker variance, to the point that a great many of the white collar speakers in the corpus pattern along with blue collar speakers. The wide distribution of white collar speakers suggests that there are more nuanced social phenomena at work within these occupational categories.
To break down the differences in the traditional white-collar category, we employ an embeddedness perspective, paying special attention to the differences between Raleigh’s traditional government jobs and newer technology-based fields. Additionally, we investigate the effects of the growing service economy on vernacularity in the service industry. In sum, we proceed with two principal hypotheses:

1. Speakers working within locally-embedded industries will exhibit more Southern shifting of vowels than other white-collar sectors.
2. Service work will show evidence of ghettoization, and speakers in the Service industry will pattern closely to blue collar speakers with respect to Southern Vowel Shift features.

3 Data and Methods

3.1 The Southern Vowel Shift and Its Decline in Raleigh, NC

Raleigh lies within the Southern dialect region (Labov, Ash and Boberg 2005), meaning that it displays elements of the Southern Vowel Shift, or SVS. The SVS is characterized by the raising and fronting of the front lax vowels (/ɪ/, /ɛ/, /æ/), the backing and lowering of the front tense vowels (/i/, /e/), and the ungliding of /aɪ/. The shift is proposed (Labov 1994) to have occurred in a three-stage process. First, /au/ monophthongizes (either categorically or before voiced consonants and codas), then the nuclei of the tense/lax /e/ and /e/ pair shift, and finally the /i/ and /ɪ/ nuclei shift. The third stage of the SVS has not moved to completion in all areas of the South, and Raleigh’s most Southern-shifted speakers only exhibit partial completion of this change. Additionally, because of the high markedness of /au/ within the Southern vowel system, we restrict our analysis to the five front vowels.
Since the establishment of RTP in the mid-1960s, Raleigh has shown a rapid retreat from the SVS (Dodsworth 2013, Dodsworth and Kohn 2012). The youngest speakers in the area exhibit a relatively mainstream vowel system, with little evidence of SVS features.

3.2 Data Coding and Modeling

This paper uses a subset of 162 speakers from the larger Raleigh Corpus of conversational interviews conducted with individuals who grew up in the Raleigh area. Interviews were transcribed and force-aligned, and tokens of the five front vowels were automatically measured at 25% of the vowel duration, corresponding to the vowel nucleus, which is the chief part of the vowel affected in the SVS. Hand correction of formant measurements were carried out where necessary. Tokens with duration under 50ms or occurring before a vowel, nasal, liquid, or glide were excluded. The resulting formant measurements were normalized within-speaker using the Lobanov (1971) formula. Each token’s position along the front diagonal of the vowel space was calculated by subtracting normalized F1 from normalized F2 (Z2-Z1) (Labov, Rosenfelder and Fruehwald 2013). Each token was also coded for preceding and following place of articulation and duration in milliseconds.

Individuals’ occupational sector, a variable we term “Labor Segment”, was coded according to a 6-way categorical schema drawn from sociological perspectives on structural embeddedness and economic changes. Coding decisions stemmed from the individual’s discussion of their job history during the interview, and most speakers went into detail about job tenure at each position and expected tasks. Individuals tend to hold a number of jobs over the course of their lifespan, but most of the speakers in this dataset saw relative stability after their earlier career trajectory. Therefore, we coded each speaker according to the job at which they held the longest tenure, with some additional weight given to a job currently held by the individual. In some situations, a speaker’s job trajectory showed a great deal of variation between labor segments, making it difficult to settle on one code. In these cases, or in cases where the speaker’s job did not fit into any
existing category, we assigned the speaker a “No Code” designation.

<table>
<thead>
<tr>
<th>Labor Segment</th>
<th>Description</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Law/Gov’t</td>
<td>Jobs with law firms or state government; historically embedded within Raleigh’s economy</td>
<td>12</td>
</tr>
<tr>
<td>Tech/FIRE</td>
<td>Technology or Finance, Insurance, and Real Estate (FIRE); new, high-growth industries</td>
<td>32</td>
</tr>
<tr>
<td>Care</td>
<td>Jobs with emotional labor and care for individuals</td>
<td>17</td>
</tr>
<tr>
<td>Service</td>
<td>Routinized interactive service work with money at the center of interaction</td>
<td>49</td>
</tr>
<tr>
<td>Professional</td>
<td>Credentialed, relatively autonomous occupations</td>
<td>16</td>
</tr>
<tr>
<td>Blue Collar</td>
<td>Work involving manual labor</td>
<td>17</td>
</tr>
<tr>
<td>No Code</td>
<td>Not classifiable with any code</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 1: Labor segment codes with number of speakers.

Linear mixed-effects regressions were run for each of the front vowels using the normalized vowel diagonal as the dependent variable. Models were constructed stepwise, and AIC comparison (Burnham and Anderson 2004) was used to assess model fit.

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Added Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Internal Factors</td>
<td>Preceding and following place of articulation, random slopes for duration by speaker, random intercept for word</td>
</tr>
<tr>
<td>Model 2: Year of Birth</td>
<td>Model 1 + Year of birth</td>
</tr>
<tr>
<td>Model 3: Year of Birth (quadratic)</td>
<td>Model 1 + Year of birth (quadratic)</td>
</tr>
<tr>
<td>Model 4: Year of Birth (cubic)</td>
<td>Model 1 + Year of birth (cubic)</td>
</tr>
<tr>
<td>Model 5: Occupation</td>
<td>Model 4 + traditional occupation coding</td>
</tr>
<tr>
<td>Model 6: Labor Segment</td>
<td>Model 4 + labor segment coding</td>
</tr>
</tbody>
</table>

Table 2: Linear mixed-effects regression models.

In addition to the incorporation of year of birth as a standard linear predictor, we test the fit of both quadratic and cubic year of birth terms, since we might expect age to have a non-linear effect in this dataset, given the pattern of community change represented in Figure 2. Furthermore, the S-curve of linguistic change within a community (Labov 2001) hypothesizes non-linearity, especially in a dataset like Raleigh’s, which includes speakers who were born before dialect contact began.

To implement the quadratic and cubic terms, we use the poly() function in R to generate orthogonal polynomials, eliminating collinearity between each year of birth term. Due to the nature of linguistic changes over time, we might also expect some differences in the rate of change in the front vowels for different labor segments in Raleigh, suggesting the inclusion of an interaction term. Interacting a polynomial with a complex categorical variable like labor segment generates a great deal of interaction terms, requiring substantially more data than we have available to accurately specify the model. Therefore, we exclude these interactions in our model testing procedures.

Lastly, we do not include sex in these models due to the high degree of correlation between sex and certain labor segments. In general, high-wage occupations tend to be male-dominated (England, Thompson and Aman 2001), and occupations with a higher proportion of women see lower overall wages as a result (England, Allison and Wu 2007). Female-dominated occupations tend to cluster within industries involving emotional labor (Dwyer 2013), meaning that care work has a strong correlation with sex. Women also tend to be locked out of high-wage positions within the managerial field (s and Ross 1992), forcing them into non-managerial roles or specialized management positions. As a result, jobs in both the professional and technology sectors would also be expected to have a high correlation with sex. Since this paper tests the effects of occupation, we choose to exclude sex as a variable from any regressions to avoid becoming mired in issues of collinearity. In addition, previous analyses of social factors affecting the SVS in Raleigh have not shown any significant effects of sex (Dodsworth 2013, Dodsworth and Kohn 2012), so it is unlikely that our analyses of occupation presented here could be reducible to sex instead.
4 Results

Table 3 shows the model fit statistics for the models run on the normalized vowel diagonal of /e/. Lower AIC signifies improved model fit, and the weighting of each model represents a relative confidence that a given model is the best fit for the data.

<table>
<thead>
<tr>
<th></th>
<th>K</th>
<th>AIC</th>
<th>AICc</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation</td>
<td>21</td>
<td>25389</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>Age Cubed</td>
<td>19</td>
<td>25397</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Labor Segment</td>
<td>25</td>
<td>25399</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Age Squared</td>
<td>18</td>
<td>25406</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>17</td>
<td>25423</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Internal Factors</td>
<td>16</td>
<td>25540</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Model fit statistics for /e/ regression models.

There are a number of interesting points to note when examining the results of the model comparisons in Table 3. First, the inclusion of polynomial age terms dramatically improves model fit for changes in the /e/ vowel, and the cubic term provides a much better fit (ΔAIC = 26) when compared with the model containing only a linear term. This degree of improvement extends to the other front vowels; in no case does a linear term outperform a polynomial. For all vowels in this dataset, the cubic age term provides either the best fit to the data or a nearly identical fit (ΔAIC < .08) to that provided by the squared term. From a theoretical perspective, we would expect changes to operate as a cubic curve, since linguistic changes that proceed through the community have stability both preceding and following a period of change. The data bear out this theory, confirming previous findings from analysis of sociolinguistic data from Philadelphia over time (Labov, Rosenfelder and Fruehwald 2013).

In contrast, the addition of labor segment shows no improvement over a traditional three-way occupation variable. The same pattern holds for all the other front vowels; the labor segment variable never outperforms occupation. This is likely related to the fact that the labor segment variable has seven categories rather than three. Even if labor segment does not provide the best model for our dataset, it may still provide insight into the differences between certain industries with respect to the retreat from the SVS. Therefore, we will explore the output from the labor segment models for each front vowel, paying special attention to the behavior of locally-embedded industries and interactive service work.

Figure 3 shows the fitted results from the linear regression model of the /e/ diagonal that includes the labor segment variable. All speakers in the corpus, regardless of occupation, show a substantial shift away from the Southern-shifted variant of /e/ over time. For specific labor segments, we are mainly concerned with the behavior of Law and Government workers (light green), Tech and FIRE workers (pink), and Service workers (purple) in relation to Blue Collar workers (red-orange). Blue Collar workers represent a vernacular baseline for the community, to which we can compare the behavior of other groups. An embeddedness perspective would predict that Tech and FIRE workers would show more deviation from Blue Collar vernacular than Law and Government workers, and our model results are consistent with this hypothesis. Workers in technology, finance, insurance, and real estate show less Southern-shifting in /e/ (t = 2.89) than the vernacular baseline. Contrastively, Law and Government shows no substantial difference (t = .98) from Blue Collar, indicating a /e/ diagonal closer to what would be expected in the SVS. In fact, Law and Government is the only Labor Segment to show little difference from Blue Collar workers, as both Care (t = 2.02) and Professional (t = 2.12) pattern closely to Technology and FIRE workers.
The model also provides evidence for the second hypothesis concerning the ghettoized nature of interactive service work. Service workers show little difference \( t = 1.50 \) from those in the blue collar industry.

The model results for both /ɪ/ and /i/, displayed in Figure 4, tell a similar story with regards to the two main hypotheses for labor segments in this dataset. Tech and FIRE workers show substantial difference from the baseline vernacular for both /ɪ/ \( (t = -3.09) \) and /i/ \( (t = 2.13) \), despite the fact that the SVS did not affect the high vowels to the same degree as the mid vowels. Law and Government workers pattern nearly identically to Blue Collar workers for both /ɪ/ \( (t = -7.4) \) and /i/ \( (t = .39) \), supporting the notion that deeply locally-embedded industries would shift away from SVS features to a lesser degree than those that have come to the area relatively recently. Likewise, the Service category patterns very closely to Blue Collar for the diagonal of /ɪ/ \( (t = -8.9) \) and /i/ \( (t = .01) \), evidencing the low status afforded to this industry.

Models of /æ/ and /ɛ/, however, show some differences from the preceding vowels. /æ/ shows no substantial occupational effects, unlike every other front vowel under study. Unlike the other four front vowels where the changes associated with the SVS are fairly unique, the raising of /æ/, which occurs in a number of other dialects (Labov 1972, 2001) may not index Southernness as strongly. The lack of a strong social connotation of Southernness in turn leads to a less concerted effort on the part of social groups to distance themselves from the raised pronunciation. /ɛ/ also shows no significant differences between occupations when compared with Blue Collar workers. However, when compared to Service workers (who are the most Southern-shifted for this vowel), Tech/FIRE once again leads \( (t = 2.26) \) in the movement away from SVS features.

Figure 3: Fitted values for Normalized /ɛ/ diagonal by year of birth and labor segment.
5 Conclusion

The results of this study illustrate the promise of incorporation of sociological perspectives on work and occupations. The linear regression models including the Labor Segment variable support our hypotheses drawn from both network embeddedness and changes in the overall structure of the labor market. Workers in technology and related industries show less evidence of Southern-shifting in the front vowels in four of five cases, supporting the hypothesis that workers in industries that recently moved into the Raleigh area would lead in changes away from the SVS. Con-
versely, workers in Law and Government never show a substantial difference from Blue Collar workers, reflecting the historical embeddedness of these jobs in the region. Lastly, Service workers pattern closely to Blue Collar workers in the front vowels, sometimes showing even more Southern patterns. Their vernacularity with regards to the SVS reflects the ghettoization of service work, both locally and within the labor market writ large.

Despite these findings in support of our hypotheses, we remain cautious about overstating the effectiveness of the labor segment schema as it stands. Without more data, the 7-way categorical variable remains too fine-grained, and it does not improve model fit over a traditional white-collar/blue-collar coding schema. The bulk of the variation due to social class or occupation exists between blue-collar speakers and other occupations, making implementation of more occupational categories a difficult endeavor. Even so, the inclusion of labor segment provides us with information we did not have before; namely, white-collar speakers variably retreat from the SVS depending on the historical embeddedness of their industry in the local labor market. This finding reinforces the importance of taking into account both the local linguistic marketplace (Bourdieu 1991) and the local labor market when incorporating social class and occupation measures into models of linguistic variation. Furthermore, these initial results highlight the promise of sociological theory and perspectives for helping linguists better analyze the social processes that underlie the correlations we find between social categories and variation.

References

England, Paula, Paul Allison and Yuxiao Wu. 2007. Does bad pay cause occupations to feminize, does feminization reduce pay, and how can we tell with longitudinal data? Social Science Research 36:1237–56.


