AMERICAN REGIONAL LEXICAL SURVEY:
GENDER AND AGE IN LEXICAL CHANGE
IN THE SOUTHERN UNITED STATES

Carol Little

Abstract
The American South has always been a distinct linguistic region. Using data from the American Regional Lexical Survey, this study shows the overall decline in use of Southern lexical terms. The following explores these changes in lexical choice in this region by comparing gender over time. Women’s choice to use Southern lexical items decreases whereas men’s usage of Southern lexical items increases significantly in the youngest generation. The results from this survey depict the effects of changing population demographics and labour statistics on choice of lexical item.

Key words
dialectology, sociolinguistics, lexical choice, regional terms, lexical survey

1 Introduction
The South has always been represented and thought of as a separate region from the rest of the United States. Some even consider it the most distinctive region in the States (Frank 1999, as quoted in Jansson 2003). The South has been portrayed as “racist, violent, poor, intolerant, and xenophobic” (Jansson 2003: 350). The roots of such a portrayal lie in the history of the American South. Though recent positive attitudes towards the South are on the rise, the South has not been able to completely shed this negative image (Jansson 2003). All references to the South hereinafter refer to the southeastern and south-central region of the United States.

Linguistically this history and portrayal of the South has several impacts. Firstly, Southern speech is stigmatized. Preston (2010) shows that Michiganders have strong negative attitudes to the correctness and pleasantness of Southern speech. In a perceptual dialectology survey where he asked Michiganders to rate the most correct English of the United States, participants always rated the South (and the New York City region) the lowest. Secondly, Southerners, aware of this stigma are linguistically insecure. Southerners also rated their own speech for correctness. The Southerners, unlike their Northern counterparts, rate their own speech relatively low for correctness but high for pleasantness. Preston (2010:
areas with greater insecurity focus on regional solidarity to express local identity”. This in turn could possibly affect lexical choice in the South, as Southerners want to express local solidarity through choice of Southern lexical items. Finally, increasing urbanization of the South (Tillery et al. 2004, Goldberg 2011) brings people from many different backgrounds together. All these people bring with them their way of speaking. This can lead to changes in certain urban environments. For example, many natives of Charlotte, North Carolina remark on the fact that it is extremely difficult to find a ‘real’ Charlotte accent (Voices of North Carolina). Some Southerners have strong views of this increasing urbanization and increasing amounts of non-native residents. In the series Voices of North Carolina, one elderly Charlottean woman bemoans:

These Northerners come down and we take them in and before you know it things change. They don’t think and act like we do. Well they sure don’t talk like us. They have a sharpness to their speech. Don’t you think so?... I feel like we [Southerners] have kind of a soft, melodic [trails off]… course why should I think it I don’t know any different. (Hutcheson 2005)

All the factors stated above have affects on the speech of the South. The American Regional Lexical Survey (ARLS) came about with this interest in investigating regional differences between the Northern and Southern United States, with an aim to determine what the current trend is. Phonetically and phonemically there exists a distinct Southern region (Labov, Ash & Boberg 2006). Lexically there is still a distinct region as well (Johnson 1996). The purpose of this study is to further investigate patterns emerging due to historical and social change as well as to determine the possible trend for the future.

2 Methodology

The survey created aimed to capture lexical use in the South. Both an online Googledoc-powered survey as well as paper copies of the survey was used as a means of dissemination to participants. Social media websites and email were used to advertise the survey. There were four demographic questions used to determine the background of the participant: the region where the participant currently lives, region(s) in which the participant has lived from four to 18 years of age, their year of birth, and their gender. The 22 lexical questions covered a range of items varying from emotive expressions to nomenclature of insects. Tokens were selected based on previous dialect dictionaries and research like the Dictionary of American Regional English (Cassidy et al. 1985), the online
Dialect Survey (Vaux & Goldner n.d.) and *Lexical Change and Variation in the Southeastern United States* (Johnson 1996). Informal discussions amongst American English speakers about different words they have heard were also taken into account. The questions covered mostly concepts that can be encountered in contemporary society. Questions included all variants of a certain item found through previous linguistic research as well as online database research. The items were in random order so as not to create a bias for the first variants.

I gathered over 200 participants but was only able to use 65 for the Northern region and 64 for the Southern region. Figure 1 represents a map of these ARLS respondents. To obtain the forms that are ‘most regional’ I only used data from those having lived in the same dialect region between the ages of four and 18 who live in the same region. Because different lexical items can be acquired more easily, unlike phonological or phonetic processes (Sankoff 2001, as quoted in Labov 2007), it was crucial to use those who have lived in the same region most of their lives. Though people having moved around during their early years are interesting in looking at lexical acquisition, it would be unclear from their lexical choices by which region they are most likely influenced. Thus, I have used only answers from people who have stayed in the same region for at least ten years of their youth and currently live in that region.

The oldest generation (born between 1920 and 1939) was treated differently because of the high mobility rate for this socioeconomic group of elderly in the United States. Many elderly go to retirement homes once they become pensioners. Due to this, many of the retirement homes they chose are not in their native region. This is the case for many of the respondents in the oldest generation. Many responses from older age groups came from a retirement home in Durham, North Carolina. Because of this, more emphasis was placed on the answers given for the dialect regions where they had lived from ages four to 18. For example, a woman who grew up in Vermont but now lives in Durham, NC (at the retirement home) is in the Northern data. Though not ideal, with the modest means and the scale of this survey such a decision was necessary in order to include participants from a vast range of ages.

The socioeconomic class and education level of survey participants are relatively similar. The subjects born in the 1990s are attending college or have plans to matriculate. Those born before 1990 have attained at least a bachelor’s degree if not higher. Thus this survey gives a snapshot of lexical variables used by the middle class with at least some university education. The networks to which I dispersed the survey were primarily of Christian European ancestry, therefore racial and ethnic background of participants can be assumed to be similar.
The survey allows the participants to select more than one variant with an extra slot to add a variant unlisted. This is advantageous because it can show regions where there are two variants coexisting. It can also be disadvantageous in that participants may select more variants than they actually use. Responses with multiple answers can be difficult to categorize when analyzing the data. For this survey I chose to use only the responses where participants chose one answer. Participants selecting multiple answers were aggregated into one category. Unless otherwise specified, I categorized regions by this analysis. However, when calculating the index of Southernness (cf. Section 2.1 below) participants selecting multiple answers are taken into account for their index scores.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>n women</th>
<th>n men</th>
<th>n total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920-1939</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>1940-1959</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>1960-1979</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>1980-</td>
<td>15</td>
<td>14</td>
<td>29</td>
</tr>
</tbody>
</table>
total      | 34      | 30    | 64      |

Table 1: The number of participants in each population group

[Map of the United States with dots indicating respondents]

Figure 1: A summary of respondents used in this survey. Each dot represents one respondent. No. of participants for the Northern region = 65. No. of participants for the South = 64. Total No. of respondents = 129

Regions were defined based on regions in the Atlas of North American English (ANAE) (Labov et al. 2006). These regions were for the most part based on phonemic and phonetic data (some lexical data was included, though it was minimal). To further investigate the South as a whole, I combined the three
Northern regions (The North, New York City, and the Mid Atlantic). Thus the Northern region acts as the region to which the South is compared. This is to discover whether the South acts as a distinct region compared to the North or if these two regions are lexically similar. This approach was taken because of the notion of the South as different than other regions of the States as explained above. Further analysis based on age and gender in South was then conducted to investigate the existence of sociolinguistic patterns. Age was separated into four categories, those born between 1920 and 1939, those born between 1940 and 1959, those born between 1960 and 1979, and those born after 1980.

Data from all questions were analyzed and tested using chi square tests. The questions that were significant (p < .05) were thus used to create the index of ‘Southernness’. Correlation within the South between age and gender was analyzed by use of the Pearson product moment correlation coefficient (r).

2.1 Index

Each participant was given an index score that determined how Southern he or she is, i.e. an index of Southernness. The more Southern terms the participant used, the more Southern he or she is on this index. The variants exhibiting a significant distribution were included in the index (p < .05). There were a total of eleven Southern terms from ten questions displaying significant distributions. One question produced two Southern variants – coke and soft drink for carbonated beverage. One point was given to each of the ten variables that had a significant distribution in the South. There were also some participants selecting more than one variable. Thirteen per cent of all the responses analyzed for the index of Southernness had more than one response. The quantity of multiple answers necessitated the use of the following equation to determine the point given: where $x = \text{point for question}$, $a = \text{southern variant(s)}$, $b = \text{total number of answer choices}$.

$$x = \frac{a}{b}$$

So for example a participant selecting you guys, you all, and y’all for their use of the second person plural would be given score of .3 for this question (all are rounded to the nearest tenth decimal). The Southern variant is y’all therefore this would be $1 ÷ 3$, which gives us .3 (I round to the nearest tenth). This method assumes that the participants choosing more than one variable use these variables in equal amounts. This assumption is problematic because for some participants
this may not be the case. Although as this survey did not take into account ratio of use for particular variables, the above method accounts for the responses where participants selected more than one answer.

3 Results

There is a small negative correlation overall of age and index score \( r = -.25 \), thus the younger generations show a decline in usage of Southern lexical items (cf. Figure 2 for summary of all index scores). Table 2 illustrates which groups have a significant distribution. The youngest group displays statistical evidence that there is a changing pattern. Reasons as to why this is the case are discussed below. When the data is separated by gender and age, other patterns emerge.

There is a fairly strong negative correlation amongst the women \( r = -.61 \): the younger the woman is the less Southern she is lexically. This trend holds fairly strongly from the oldest group to the youngest group of women the women (cf. Figure 4).

Amongst the men, the correlation is not so clear. There is little correlation overall in the men born from 1920 onwards \( r = .12 \) (Figure 5). Once the data is separated there is a very strong positive correlation of the men born between 1920 and 1979 \( r = .89 \) (Figure 6). As men get younger within this age group their Southernness increases. When the two youngest generations are compared (those born in 1960 and onwards), there is a fairly strong negative correlation \( r = -.64 \) (Figure 7). For the last two generations as age decreases, so does index of Southernness, thus use of Southern variables is decreasing amongst the youngest generation of men (cf. Table 1 for numbers in each population group).
Figure 2: The index scores and birth years of all participants from the south (n=64). Each polygon represents one participant. Cf. Table 1 for population sizes. \( r = -0.25 \)

<table>
<thead>
<tr>
<th></th>
<th>1920-1939</th>
<th>1940-1959</th>
<th>1960-1979</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920-1939</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1940-1959</td>
<td>0.014</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1960-1979</td>
<td>0.058</td>
<td>0.463</td>
<td>-</td>
</tr>
<tr>
<td>1980-</td>
<td>0.188</td>
<td>0</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Table 2: The p values for comparisons of all age groups in the South
Figure 3: The average index score of the four age groups for men and women. Each point represents the average for one age group. Cf. Table 1 for n values for each population. Cf. Figures 4 and 5 for trends for women and men respectively.

Overall there was no significance between genders (p = .2), but there were significant distinctions between ages and genders (cf. Table 2, 3 and 4). P values in red are significant (p < .05). All are rounded to the nearest thousandth. M = men and W = women. The reasons behind these differences will be further discussed below.

Figures 4, 5, 6, and 7 depict visual representations of the trends observed.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1920-1939</td>
<td>0.003</td>
<td>0.144</td>
<td>0.108</td>
<td>0.011</td>
</tr>
<tr>
<td>W 1940-1959</td>
<td>0</td>
<td>0.041</td>
<td>0.046</td>
<td>0</td>
</tr>
<tr>
<td>W 1960-1979</td>
<td>0.101</td>
<td>0.199</td>
<td>0.017</td>
<td>0.333</td>
</tr>
<tr>
<td>W 1980-1990</td>
<td>0.215</td>
<td>0.007</td>
<td>0</td>
<td>0.333</td>
</tr>
</tbody>
</table>

Table 3: The p values for men and women age groups according to their index scores

Table 4: The p values for the men age groups comparing index scores.
Table 5: The p values for the women age group comparing index scores

<table>
<thead>
<tr>
<th></th>
<th>W 1920-1939</th>
<th>W 1940-1959</th>
<th>W 1960-1979</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1920-1939</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>W 1940-1959</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>W 1960-1979</td>
<td>0</td>
<td>0.005</td>
<td>-</td>
</tr>
<tr>
<td>W 1980-</td>
<td>0</td>
<td>0</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Figure 4: The index score for women with a trend line (in black). Each dot represents one respondent. $r = -.61$

Figure 5: The index score for men with a trend line (in black). Each dot represents one respondent. $r = .12$
Figure 6: The index score of men born between 1920 and 1979 with a trend line (in black). 
$r = .89$

Figure 7: The index scores for men born in 1960 and after with a trend line (in black). Each dot represents one respondent. $r = -.64$
4 Discussion

The results in the lexical data somewhat conform to other research done on phonological features on the Southern Shift, a phonological process affecting vowels in the dialect of the Southern United States. In other studies, older generations still show an advanced progress of the Southern Shift although this shift seems to be receding in the youngest generation, with the women being the leaders of this new recession of Southern phonological patterns (Jacewicz et al. 2011). Fridland (2006) has also found similar trends in Memphis where the youngest generation is not displaying certain aspects of the Southern Shift that seemed to be progressing among the older generations. In this study as well the women are leading in this change. Fridland does comment how apparent time data and gender differences do portray a somewhat complex relationship. Lexical data from the Southeastern United States concludes that the gap in men and women’s lexicons have not changed between 1930 and 1990 (Johnson 1996). The data from ARLS present a different pattern conforming more to the phonological and phonemic data of Jacewicz et al. (2011) and Fridland (2006) rather than the lexical data presented by Johnson (1996).

I will discuss first an analysis as to why such different trends emerge in the different gender groups and I will then tie this into the greater picture for the South and the future of Southerners’ lexical choices.

The image of the woman in the South is complicated: her portrayal even in the late twentieth century reflects relics of how a Southern woman should be. In
1979, a man thought the best choice for a companion reflecting “old-fashioned, Christian virtue… a lady in the finest sense” was a Southern woman (Rogers 1982: 60). These data point to a change in Southern women. There are several factors that could influence the decrease of Southern orientation for women within the generations studied. This can be attributed to the increase of women in the workforce starting with the economic boom after WWII (De Hart 1997) as well as increasing participation of women in the labour force amongst younger generations (Juhn & Potter 2006).

Studies have represented the negative attitudes to Southern linguistic features (Preston 2010). The prestige variants are not those of the South, but the standard variants of the North. Because of the woman’s changing place in society and her increased participation in the work force, her need to assert her power is more salient. Eckert (1989: 250) argues that it is power that is the “most appropriate underlying sociological concept for the analysis of gender-based linguistic variation”. This need for assertion of power and in turn for women to express membership to prestigious powerful groups leads to the use of more prestige variants in speech. “In Western society this is perhaps most clearly illustrated in the greater emphasis on femininity in the South, where regional economic history has domesticized women and denied them economic power to a greater degree than it has in the industrial North” (Fox-Genovese 1988, as quoted in Eckert 1989: 257). Because powerlessness is attributed to women, women in the South especially need to claim status now that they are occupying a growing place in the job market. Nichols (1983) explains that access to jobs and education amongst women contributes to their use of the standard forms. Therefore I argue that because of increase in labour participation, women are using more standard (and thus overtly prestigious) variants to manifest symbolic connection to the more prestigious class.

Men born between 1920 and 1979, on the other hand, present a different pattern than women. Their index scores have a positive correlation: the younger the man the more Southern he is in the oldest three generations (cf. Figure 6). During this same time period (1920-1990) the influx of people born outside the Southern region increases. Bailey et al. (1993) shows how increases of populations could amplify the spread of regional terms, as the residents want to express solidarity to the region by emphasizing regional speech. The men born between 1920 and 1979 exemplify this in their use of more Southern lexical variants in order to assert solidarity and express their strong ties to the local community. For the men in the first three generations, their notion of power lie in solidarity of region, thus they express this by using more Southern lexical variants.
Yet, to account for the sudden drop in Southernness in the youngest generation of men \((r = -0.64)\), I look again to labour statistics. As aforementioned, women are participating more in the workforce. A current trend in the United States is the increasing number of wives who earn more than their husbands. In 1987, 24 per cent of wives earned more than their husbands, where in 2006, 33 per cent of wives earned more than their husbands. This figure in 2006 is around a third higher than the figure in 1987 (Bureau of Labour Statistics 2009). Women’s earnings growth is also higher at all education levels between 1979 and 2000. Especially pertinent to this survey is that “earnings for women with college degrees have increased by 30.4% since 1979 on an inflation adjusted basis, while those of male college degree graduates rose only by 16.7%” (Bureau of Labour Statistics 2001). One of the most salient and most recent trends that could be affecting lexical choice among men is that in 2008 women who were single and childless between the ages of 22 and 30 earned 8 per cent more than their male counterparts in most cities in the United States (Dougherty 2010). This suggests an impending threat to men’s power in the workplace. The labour force used to be overwhelmingly dominated by men, especially in the South where the traditional image of an old-fashioned Southern housewife was valued. The sudden drop of average index score between men born after 1960 exhibits a fairly strong negative correlation \((r = -0.64)\). Therefore I argue that men in the youngest generation are using more symbolic methods of expressing prestige as the threat of women earning more than men becomes evermore increasing.

However, Eckert (1989: 55) shows that “although employment conditions may change, the underlying relations of power and status between men and women can remain quite unchanging”. She continues to explain, “actual power relations between men and women can be expected to lag behind”. Since lexical items are most readily acquirable (Sankoff 2001, as quoted in Labov 2007), changes in lexicons over several generations are more prone to illustrating shifts in social patterns like the demographic and labour patterns discussed above. Thus lexical items, such as in ARLS, can display more susceptibility to change than phonemic aspects. Therefore the results displayed above show this rapid change in the South amongst Southern men as a product of men’s need to assert their symbolic power as women’s power in the economy grows.

What of the larger picture? What is the fate of lexical choice overall in the South? There is a slight negative correlation between age and index score \((r = -0.25)\) suggesting that display of Southernness through lexical choice is decreasing. In the youngest generation there is a stronger correlation \((r = -0.42)\) and statistical testing signifies that between the second youngest and youngest generation this decrease is significant \((p = .01)\).
Higher mobility in American society leads to more dialogue between people of different dialect regions and in turn the exchange of lexical words. This could entail the loss of certain Southern variants in favour of standard Northern variants, as is shown amongst the women. This is attributed mainly to increased interregional dialogue rather than media as it has been manifested that TV and other forms of media do not make people all sound the same (Chambers 1998).

As there were many variants used for the index, two examples are discussed below that exemplify the changing South. One variant that did have a significant distribution in the South was the slang term *plumb*, for ‘very’. *Plumb* is virtually extinct; no participants in the youngest generation selected this as a lexical item they would use. There is also competition within the youngest generation between *lightning bug* (which was selected unanimously by all three older generations) and *firefly* to describe a bug with luminous organs, the Northern variant. Figure 8 exemplifies this decrease in the term *lightning bug* among the youngest generation and a turn towards the usage of *firefly* (a 4 x 2 chi square test confirms this with a chi square value of 17.6, thus p < .005). The number of respondents selecting both terms also increased. This could be tantamount to a shift towards the Northern variant *firefly*. *Firefly* is the term most widely used in the North and the West. It is important to note that websites, like for example *National Geographic*, describing the insect almost always give *lightning bug* as an alternative term, showing that this term is still prevalent in society. This term will most likely continue to be in use in the South but its fate has yet to be determined.

5 Conclusion

Because this survey overrepresented variants in the South, it may have not captured significant Northern variation. For the future, research on variation in the North and possible trends there would need to be conducted. Studies shed light that some Southern variants like *y’all* (Maynor 2000, Tillery et al. 2000) are starting to become prevalent in regions other than the South. It would be interesting to further investigate the diffusion of this and other Southern variants into the North and the West.

Because of the relatively small size for each regional group, both under 100 respondents, more respondents for future surveys would be beneficial to this field of research.

As exemplified in Figure 1 above, there is a high concentration of respondents from North Carolina. Thus this survey may not be entirely representative of the South as a whole. The social classes of the respondents are also limited as explained in the methodology section. Further research on this is needed.
As this survey did not capture racial or socioeconomic demographic information, for the future the inclusion of these would be beneficial in order to investigate certain trends regarding these factors. People of Hispanic origin are the most salient ethnic group in terms of contribution to population group. A survey that included such factors could determine the possible trend for different ethnic groups and thus the overall trend. In North Carolina especially there is a large Hispanic immigrant population. Research has shown that the sons and daughters of the first generation immigrants are displaying Southern tendencies lexically and phonetically (Wolfram et al. 2004). As ethnic populations increase their contribution to linguistic activity they become a more salient factor in linguistic change. As this survey displays results from participants of European descent, further investigation of this is needed.

As demographics change in this world of increasing mobility, it is difficult to determine what will happen to overall variation in regional lexicons. This survey indicates that the trend for Southerners’ lexical choice is becoming less Southern. Will increased mobility contribute to lexicon levelling in the United States? More investigation is needed to confirm this.

These data demonstrate how different social variables can affect linguistic change. It also conveys how the variables affecting change can change themselves. They reveal how certain social changes can become stronger than others and thus override the previous change in progress. Further research is needed to see if such patterns exist amongst different linguistic aspects as well a in a more varied population selection.

References