



Investigating the impact of dialect prestige on lexical decision

Mairym Lloréns Monteserín¹, Jason Zevin²

¹ University of Southern California, Linguistics

² University of Southern California, Psychology

llorensm@usc.edu, zevin@usc.edu

Abstract

The speech signal encodes both a talker's message and indexical information about a talker's identity. Dialectal variation is one way in which non-linguistic information about a talker is conveyed through her speech. A talker's dialect tends to correlate strongly with her demographic background, and listeners are known to form beliefs about speakers based on their dialect alone: talkers of lower-status dialects are consistently downgraded on positively-valued attributes relative to talkers of canonical dialects. Hypothesizing that pre-formed beliefs about a low-status talker might impact optimal perception of her speech, this study investigated the influence of the relative prestige of talker dialect on listeners' behavior in three lexical decision experiments. The finding of significantly increased propensity to incorrectly reject words uttered in an arguably low-prestige variety of American English relative to both normative General American English and British English suggests that talker status may play a role in the success with which talker messages are perceived by listeners. These results as well as unexpected interactions of dialect and word frequency in some but not all experiments are discussed in the context of signal detection theory.

Index Terms: dialectal variation, talker identity, lexical decision.

1. Introduction

Speech encodes both a talker's message and information about the talker. Types of indexical information listeners are sensitive to include fundamental frequency of an utterance, which informs listeners about a talker's relative size, age, gender and ethnicity; and dialectal variation, which informs listeners about a talker's regional/geographic background. To illustrate, listeners have been shown to use a talker's production of sociolinguistic variables and their own implicit knowledge of dialectal variation to infer the talker's geographical background [1] and race [2].

Studies looking at the influence of speaker on perception highlight aspects of processing that likely *facilitate* comprehension. For instance, shifting phoneme boundaries in response to talker categories such as gender [3], sexual orientation [4] or age [5] may *facilitate* lexical access for listeners hearing the speech of talkers that differ along these dimensions.

It is also possible that processing of indexical information *obstructs* transmission of a talker's message in some cases. For example, an American English speaker's relative social status—as indicated by her variety of American English—may affect the degree to which listeners attend to her speech, thereby impacting comprehension. Status refers to an

individual's or subgroup's standing relative to other individuals or subgroups within a larger social whole. It is well known that members of subgroups within the larger social whole enjoy different levels of prestige/stigma, where prestige and stigma can be thought of as the extremes along a continuum of high and low status, respectively.

Giles [6] reviews previous research that has looked at the effect of dialectal status on beliefs about the attributes of speakers. This research has shown that regional varieties of languages such as American English are rated as *lower-status* and more “incorrect” relative to the “standard” variety, General American English (GAE) [7]. As a result, speakers of non-standard varieties are downgraded on a number of status attributes. For instance, Luhman [8] found that listeners rated Appalachian English speakers significantly lower than GAE speakers on status attributes such as “wealthy”, “educated”, “successful”, “intelligent”, “attractive” and “ambitious”. Heaton and Nygaard [9] replicated these results for speakers of Southern English (SE) relative to GAE. Labov and colleagues [10] found that listeners evaluated speakers negatively on scales of “professionalism” as a function of the speakers' use of the [ɪn] variant of the “-ing” morpheme. Similarly, Loudermilk et al. [11] found strong associations between canonical [ɪŋ] pronunciation and “intelligence” and “high socio-economic status” attributes using an implicit awareness test.

Reviews of this research [12–13] suggest that speakers of “non-standard” varieties of American English may be considered less competent, less suitable for employment, and more likely to be considered guilty when under suspicion of having committed a crime than speakers of the “standard” variety. The question of whether or not these *attitudes toward* and *valuations of* speakers impinge on processing of their actual speech has not been investigated, however. In the current study we make a preliminary attempt to address this question by probing listeners' behavior in lexical decision (LD).

Traditionally, processing of a talker's message and identity are thought to occur independently. In this view, a high-order event like LD—which is expected to occur after phonetic feature extraction and lexical access have taken place—should not be influenced by relative talker prestige or any other indexical parameters. In contrast to this view, we hypothesized that listeners' beliefs about a talker as indexed by her dialect may influence processing of her speech by boosting or damping the attention paid to her utterances. Given that non-standard varieties of English such as SE are nevertheless intelligible to the average listener [14], we reasoned that differences in behavior during LD as a function of dialect may reflect interdependent processing of talker message and indexical information about the talker and not differences in a listener's relative unfamiliarity with non-standard varieties.

We envisioned two ways in which talker relative status could come to influence listeners' behavior during LD. On the one hand, listeners could adopt more liberal decision criteria for perceived higher-status talkers: the *benefit of a doubt* scenario. On the other hand, listeners could adopt a stricter decision criterion for lower-status talkers or tend to incorrectly reject words their words: the *less of a chance* scenario. These scenarios are not mutually exclusive.

We conceptualized the LD task as a yes/no decision amenable to analysis in signal detection theory (SDT). SDT has been used widely in LD research previously [15] and enables us to dissociate a respondent's sensitivity—her ability to discriminate between signal and noise trials—from her bias—the inherent probability that she will say yes based on some unknown decision criterion. There are four possible outcomes of an LD task. If the trial is a word trial, *yes* responses constitute “hits” and *no* responses constitute “misses”. If the trial is a nonword trial, *yes* responses constitute “false alarms” (FAs) and *no* responses constitute “correct rejections” (CRs). For calculation of sensitivity (d') and bias (β) in SDT, only *yes* responses are considered, both in signal trials when yes is the correct response (i.e., hits) and in noise trials when yes is the incorrect response (i.e., FAs).

Viability of the scenarios described above can be evaluated by looking at criterion shifts (i.e., significant shifts in average β values as a function of dialect), but one way to specifically evaluate the second scenario independently of the first is by comparing the proportion of misses within each dialect condition, which are not taken into account in calculations of sensitivity and bias. If indexical information about talkers like their social status impinge on the outcomes of LD behaviors, then we could expect to see either criterion shifts or differences in the rate of misses (or both), with shifts occurring in directions that “favor” speakers of relatively prestigious dialects. Otherwise, no such shifts should occur, or they should occur in directions that “favor” the familiar GAE relative to the unfamiliar British and Southern English.

2. Experiments

Three separate experiments were implemented. Experiment1 compared LD as a function of relative dialect prestige between normative—and therefore higher status—GAE and relatively lower status SE. As we said previously, speakers of SE are consistently downgraded relative to speakers of GAE on positively-valued attributes. Experiment2 compared LD between GAE and British English. Given that differences in LD behavior as a function of dialect in Experiment1 could feasibly be result from differences in *familiarity* with the non-standard varieties—and not due to differences in relative prestige—Experiment2 pits GAE to what is arguably an English variety that matches SE in unfamiliarity, but whose speakers are not downgraded on positively-valued attributes. Experiment3 pitted SE against British English.

2.1. Methods

2.1.1. Participants

A total of 60 participants were recruited from the University of Southern California community for participation ($n=20$ for each of Experiments1-3). All participants were either part of the Psychology Department's undergraduate subject pool and received partial course credit for their participation or were enrolled in one of two undergraduate Linguistics Department

courses and received course credit and/or extra credit for participating.

2.1.2. Stimuli

One hundred and twenty stimulus items were recorded in three distinct varieties of English for a total of three hundred and sixty recordings. The 120-item stimuli list consisted of sixty phonotactically-licit nonwords, thirty high frequency words and thirty low frequency words.

The sixty words of English were selected after consulting the *English Lexicon Project* [16] database, which provides behavioral data from lexical decisions and lexical characteristics of words and nonwords. The high frequency word list consisted of nouns with the highest “hyperspace analogue to language frequency” [17] in the database that also had zero phonological neighbors. Items with less than 0.88 mean accuracy scores in LD tasks were excluded. The low frequency word list consisted of nouns and adjectives among the lowest in HAL frequency that also had zero phonological neighbors. Items with less than 0.75 mean accuracy scores were excluded. Nonwords were constructed by first taking high frequency nouns/adjectives with three to six phonological neighbors and then altering a single phoneme. Single-phoneme alterations that would have resulted in real words were excluded.

A single, female speaker recorded the stimuli in the GAE, SE and British varieties of English. In the case of the nonwords, the speaker was provided the original word from which the nonword was derived. She was instructed to model her pronunciation of each nonword on the dialect-specific pronunciation of the original real word from which that particular nonword was derived. The speaker was selected based on her documented professional expertise in voice acting, dialect production and dialect coaching for theater, radio and film. She reported being able to produce speech in the three English varieties with ease. She read the lists of items twice in different random orders. She was instructed to read each list in GAE, defined as “newscaster English / Standard English”; British English, defined simply as “received pronunciation of British English”; and SE, defined as “Southern drawl / typical of a speaker from northeastern Texas”. The latter instruction was included in order to specify a variant within the heterogeneous Southern group of dialects of American English.

Recordings were segmented and labeled automatically in Praat [18] and verified manually. Following segmentation, all sound files were normalized for intensity, also on Praat. Next, sound files were grouped by dialect and a file of long-term average spectrum (LTAS) noise was generated for each dialect group. Finally, all sound files within a dialect group were mixed with their respective LTAS noise file on Matlab.

2.1.3. Procedure

Three versions of the same experiment design were implemented (Experiments1-3). These differed only with respect to the combination of dialects of presented stimuli. During each experiment, participants heard half of all items in one dialect and half in another. Lists of stimuli were counterbalanced. Within each dialect, half (30) of items were words and half were nonwords. Within the list of real words, half had high and half had low frequency. The one hundred and twenty trials were presented in a single, randomized block. The experiments were built and run using the psycholinguistic

experimentation software platform Paradigm [19]. Participants sat in sound-attenuated booths facing computer monitors, wearing dynamic stereo headphones. Before beginning, they were instructed to use the mouse to click on Yes/No buttons in the center of the screen after hearing each stimulus item to indicate their decision for whether or not the presented item was a word of English.

2.2. Analyses

In order to determine that participants' ability to discriminate words from nonwords remained unobstructed across dialectal conditions, subject sensitivity was calculated as a function of dialect in all experiments. To answer the question of whether or not subjects shifted their decision criteria as a function of dialect subject bias was also calculated, as was the propensity to incorrectly reject real words.

2.2.1. Sensitivity and Bias

As reviewed by Stanislaw and Todorow [20], calculation of sensitivity and bias consists of computing a value known as d' and β for each participant. Sensitivity (d') and bias (β) parameters were computed for individual subjects in each experiment using the R statistical software [21] with R code taken from Pallier [22]. d' values greater than zero indicate increasingly greater ability to discriminate words and non-words. β values around one indicate "no preference"—participants are just as likely to respond yes as they are to respond no. β values greater than one indicate a "stricter" decision criterion, or a propensity to respond no. β values less than one indicate a more "liberal" decision criterion, or a propensity to respond yes.

Group sensitivity and bias was calculated within each experiment by averaging subjects' sensitivity and bias values in each dialect condition. Mean group values were subjected to pairwise t-tests. P values less than .5 were considered significant.

2.2.2. Propensity to say "no"

Recall that we envisioned two ways that a talker's relative social status as indexed by dialect could influence listeners' LD: the *benefit of a doubt* and the *less of a chance* scenarios. Support for the former can be found by measuring any criterion shifts as part of SDT analysis. Incorectly rejected word trials may constitute a relatively independent measure of the viability of the latter scenario for each participant because they constitute trials where some factor—perhaps beliefs about attributes of the speaker based on her dialect—lead the participant to incorrectly reject an actual word. We therefore calculated the average misses for each experimental group as a function of dialect after tallying each participant's individual value and subjected the means to pair-wise t-tests. P values less than .5 were considered significant.

2.3. Results

2.3.1. Propensity to incorrectly say "no"

Dialect significantly influenced the number of misses—real word trials that were incorrectly rejected—in Experiments1 and 3, the two experiments where subjects were exposed to the SE dialect. In Experiment1, subjects missed significantly more real words in SE than in GAE, $t(39)=6.12$, $p<0.5$. Subjects in Experiment3 had a similar propensity to miss words in SE

relative to BRP, $t(39)=1.4$, $p<0.5$. Subject responses in Experiment2, which opposed GAE to BRP, showed no significant effect of dialect on the propensity to miss word trials.

Table 1. Average number of misses/standard deviation in each experiment.

Experiment	GAE Misses	SE Misses	BRP Misses	Significant Difference
1	4.45	6.3	-	✓
2	5.25	-	5.45	
3	-	8.9	5.5	✓

2.3.2. Bias

Participants in Experiment1 displayed an overall liberal bias regardless of dialect. Decision criterion shifted significantly in the liberal direction for GAE relative to SE utterances but only when considering high frequency items, $t(39)=7.24$, $p<0.5$. Bias shifted in the same direction for low frequency items but the shift was not significant, $t(39)=-0.35$, $p>0.5$.

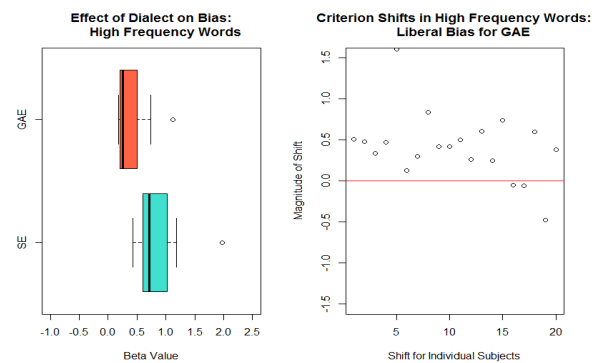


Figure 1: Frequency effect on average bias as a function of dialect in **Experiment1**. Left: Boxplots of mean bias by dialect in high-frequency words. Right: Criterion shifts in Experiment1 per subject. Distance from zero line indicates magnitude of shift. Points above zero line represent participants who shifted toward a more liberal bias for GAE relative to SE; points below represent participants who shifted toward a more liberal bias for SE relative to GAE.

Responses in Experiment2 also showed a significant shift in the liberal direction for GAE, this time relative to British English and again only among high frequency words, $t(39)=5.47$, $p<0.5$. Bias for low frequency items shifted in the opposite direction—more liberal for British English relative to GAE but not significantly so, $t(39)=1.97$, $p>0.5$. As in Experiment1, responses were liberal overall regardless of dialect.

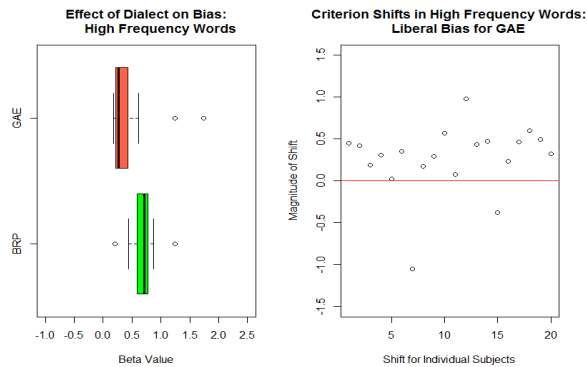


Figure 2: Frequency effect on average bias as a function of dialect in **Experiment 2**. *Left*: Boxplots of mean bias by dialect in high frequency words. *Right*: Criterion shifts in Experiment 2 per subject. Distance from zero line indicates magnitude of shift. Points above zero line represent participants who shifted toward a more liberal bias for GAE relative to British English; points below represent participants who shifted toward a more liberal bias for British English relative to GAE.

Experiment 3 showed an average bias for British English hovering just below 1, indicating a slightly liberal bias; responses to SE, however, displayed a stricter decision criterion (average bias greater than 1). Criterion shifts reflected lowered bias for British English relative to SE in both high and low frequency words; among high frequency words the shift was significant, $t(39)=1.13$, $p<0.5$ and among low frequency words the shift was not significant, $t(39)=0.22$, $p>0.5$.

2.3.3. Sensitivity

Results from Experiments 1-3 show that subjects are generally capable of discriminating words from nonwords in all three dialects—that is, average d' values were greater than zero for all three experiments ($1 < d' < 1.5$). Sensitivity differed as a function of dialect in all three experiments. Subject responses in Experiment 1 were significantly more sensitive in GAE relative to SE for both high ($t(39)=-8.43$, $p<0.5$) and low ($t(39)=-2.39$, $p<0.5$) frequency words. In Experiment 2, responses were more sensitive in GAE relative to BRP across all words but the difference was only significant within high ($t(39)=-4.28$, $p<0.5$) and not low ($t(39)=-1.69$, $p>0.5$) frequency items. In contrast, neither high ($t(39)=-0.64$, $p>0.5$) nor low ($t(39)=-0.58$, $p>0.5$) frequency items showed a significant effect of dialect on sensitivity in Experiment 3.

3. Discussion

Data on misses from Experiments 1-3 provide preliminary support for the *less of a chance scenario*—listeners were significantly more prone to incorrectly reject true words when uttered in the low-prestige SE dialect than when uttered in the relatively high-prestige GAE and British dialects. Good overall sensitivity and the lack of significant differences in average misses between British English and GAE suggests that listeners' unfamiliarity with non-standard varieties may not account for the increased rate of misses in SE relative to other dialects. The *benefit of a doubt scenario* found less

support. While there was some effect of dialect on decision criteria in all experiments, this effect was confounded by an unexpected effect of word frequency.

One explanation for word frequency effects on response sensitivity and bias takes the relatively high familiarity of GAE into account. If LD occurs in a two-step process as suggested by Balota and Chumbley [23], then a fast-response mechanism may be reserved for high frequency words uttered in the ubiquitous GAE dialect, with a slow-response analytical decision stage occurring only if the fast-response mechanism fails. To put another way, in experiments where GAE was heard, the tendency for responses to “favor” GAE in high frequency words may have reflected the extent to which these particularly common words are familiar to listeners, specifically when uttered in the normative and ubiquitous GAE. Listeners are unfamiliar with low frequency words, regardless of dialect, so these items might have been relegated to the slower analytical stage in all experiments. Criterion shifted in favor of the arguably higher-status British English in Experiment 3 with sensitivity remaining constant, so though frequency effects confuse interpretation of our study, results are largely consistent with some role for relative dialect status.

4. Conclusions

The findings reported here on incorrect rejections across the three studies and criterion shifts in Experiment 3 lend some support to the notion that talker identity and message are not processed independently but together constitute a coherent signal. At what point in processing these two signal components become integrated is a question for further research. One suggestion from this study is that listeners may modulate the attention they pay to speech of talkers of different dialects based on pre-formed beliefs about those talkers. Attention could be modulated in terms of “quality” or in terms of “time paying attention”. To test if listeners pay less attention to utterances by low-prestige talkers, thereby missing crucial cues for lexical access and other information required for comprehension, reaction times for LD responses in these studies could be analyzed.

Further experimentation is needed to pinpoint the role of relative talker/dialect status in comprehension. A major shortcoming of this study is its lack of ecological validity. Experimental designs looking at more naturalistic behaviors than LD are likely to shed light on the time-course and functioning of mechanisms that integrate—and keep apart—processing of a talker's identity and message. Continuing to probe LD, utterances by native speakers of each of the dialects might be more appropriate stimuli to capture naturalistic listener behavior. Despite the limitations of this study, converging evidence suggests that the role that talker information plays in perceptual outcomes should be further studied.

5. Acknowledgements

Funding was provided by a Graduate Fellowship from the Dornsife School of Arts and Letters and NIH grant T32 DC009975. The first author thanks Brian Baumann and students at Dr. Zevin's Cognitive Neuroscience lab without whom this project would have been impossible.

6. References

- [1] C. Clopper, S.V. Levi and D.B. Pisoni, "Perceptual similarity of regional dialects of American English," *JASA*, vol.119, pp. 566-574, Jan. 2006.
- [2] L. Staum, "Experimental investigations of sociolinguistic knowledge", Ph.D. dissertation, Dept. of Linguistics, Stanford U., Stanford, CA, 2008.
- [3] E. Strand, "Uncovering the role of gender stereotypes in speech perception," *J. of Language and Social Psychology*, vol.18, no.1, 1999.
- [4] B. Munson, "Levels of phonological abstraction and knowledge of socially motivated speech-sound variation: a review, a proposal and a commentary on papers by Clopper, Pierrehumbert and Tamati; Drager; Foulkes; Mack; and Smith, Hall and Munson," *J. of Laboratory Phonology*, vol.1, pp.157-177, 2010.
- [5] J. Hay, P. Warren and K. Drager, "Factors influencing perception in the context of a merger-in-progress," *J. of Phonetics*, vol.34, pp.458-484, 2006.
- [6] H. Giles, "Language attitudes: the role of information management," *Canadian Issues*, pp.28-34, August 2011.
- [7] T.C. Frazer, "Attitudes toward regional variation," *J. of English Linguistics*, pp.89-100, 1987.
- [8] R. Luhman, "Appalachian English stereotypes: language attitudes in Kentucky," *Language in Society*, vol.19m no. 3, pp.331-348, September 1990.
- [9] H. Heaton and L. Nygaard, "Charm or harm: effect of passage content on listener attitudes toward American English accents," *J. of Language and Social Psychology*, vol.30, no.2, pp.202-211, 2011.
- [10] W. Labov et al., "Properties of the sociolinguistic monitor," *J. of Sociolinguistics*, vol.15, no.4, pp.431-463, 2011.
- [11] B.C. Loudermilk, "Cognitive mechanisms in the perception of sociolinguistic variation," PhD diss., University of California Davis, 2013.
- [12] H. Giles and A. Billings, "Language attitudes," *Handbook of applied linguistics*, Oxford, UK: Blackwell, pp.187-209, 2004.
- [13] H. Giles and M. Marlow, "Theorizing language attitudes: past frameworks, an integrative model and new directions," *Communication Yearbook*, Thousand Oaks: Sage, pp.161-197, 2011.
- [14] C. Clopper and A. Bradlow, "Perception of dialect variation in noise: intelligibility and classification," *Language and Speech*, vol.51, no.3, pp.175-198, 2008.
- [15] A. Jacobs, R. Graf and A. Kinder, "Receiver operating characteristics in the lexical decision task: evidence for a simple signal-detection process simulated by the multiple read-out model," *J. of Experimental Psychology*, vol.29, no.3, pp.481-488, 2003.
- [16] D.A. Balota et al., "The English lexicon Project," *Behavior Research Methods*, vol.39, pp.445-459, 2007.
- [17] K. Lund and C. Burgess, "Producing high-dimensional semantic spaces from lexical co-occurrence," *Behavioral Research Methods, Instruments and Computers*, vol.28, no.2, pp.203-208, 1996.
- [18] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Computer Program, Version 6.0.15.
- [19] Perception Research Systems, "Paradigm stimulus presentation," Computer Program, Web, 2007.
- [20] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behavioral Research Methods, Instruments and Computers*, vol.31, no.1, pp.137-149, 1999.
- [21] R Core Team, "R: A language and environment for statistical computing," *R Foundation for Statistical Computing*, Vienna, Austria, URL <http://www.R-project.org/>, 2015.
- [22] C. Pallier, "Computing discriminability and bias with the R software," URL <http://pallier.org/resources/aprime/aprime.pdf>, 2014.
- [23] D. Balota and J.I. Chumbley, "Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage", *J. of Experimental Psychology: Human Perception and Performance*, vol.10,no.3, 1984.