

Received 14 July 2017.

Accepted 3 October 2017.

SURVEYING BORDERS IN A SPEECH COMMUNITY

Christopher STRELLUF & Tasha CARDWELL

Northwest Missouri State University*

cstrell@nwmissouri.edu / tashabcardwell@gmail.com

Abstract

This research reports responses to a language questionnaire for an extensive set of phonological, lexical, and grammatical variables in Kansas City. Linguistic variables are modeled with conditional inference trees and random forests against demographic factors like age, race, and sex, as well as several geopolitical factors. Results show that—while Kansas Citians identify many cultural, political, and social barriers that shape life in the city—race is the primary social predictor of survey responses. Among sixteen survey items with significant predictors, race is most significant for nine. Geopolitical factors that Kansas Citians imagine as meaningful do not affect linguistic practice. Based on this finding, the article considers the “speech community” as a unit of sociolinguistic analysis and concludes that the large metropolitan area can be thought of as a single speech community, as long as the sample is racially homogenous.

Keywords

questionnaire, borders, race, speech community, Kansas City

EXAMINANDO FRONTERAS EN UNA COMUNIDAD DE HABLA

Resumen

Esta investigación da cuenta de las respuestas a un cuestionario compuesto de un conjunto extenso de variables fonológicas, léxicas y gramaticales preguntado en Kansas City. Las variables lingüísticas se modelaron a partir de árboles de inferencia condicional y de bosques aleatorios aplicados a

* Author 1: S1.74 Social Sciences Building, University of Warwick, Coventry CV4 7AL, United Kingdom /
Author 2: 13785 Hunter Drive, Platte City, Missouri 64079, United States

factores demográficos como la edad, la raza y el sexo, así como a varios factores geopolíticos. Los resultados muestran que, si bien los ciudadanos de Kansas identifican muchas barreras culturales, políticas y sociales que configuran la vida en la ciudad, la raza es el principal predictor social de las respuestas de la encuesta. Entre los dieciséis ítems de la encuesta con factores predictivos significativos, la raza es más significativa en nueve. Los factores geopolíticos que los ciudadanos de Kansas creen que son significativos no afectan a la práctica lingüística. Sobre la base de este hallazgo, el artículo considera la “comunidad de habla” como una unidad de análisis sociolingüístico y concluye que la gran área metropolitana puede considerarse como una sola comunidad de habla, siempre que la muestra sea racialmente homogénea.

Palabras clave

cuestionario, fronteras, raza, comunidad de habla, Kansas City

1. Introduction

This paper reports responses to a language questionnaire administered in Kansas City, a large metropolitan area in the US Midwest. Data are drawn from the Missouri Language Survey (MLS), an instrument originally developed and written by Matthew J. Gordon and described in Gordon’s (2006) study on the low back vowel merger in Missouri. The present study makes use of a slightly modified version of the survey.

We present this research with three inter-related goals. First, we wish to document several phonological, lexical, and grammatical features that have been identified in the region, but not systematically studied in Kansas City. Second, we wish to follow up on previous sociolinguistic fieldwork in Kansas City, which identified several salient geographical, social, and political borders that may affect residents’ linguistic practices. Third, we wish to examine the extent to which a large metropolitan area can be thought of as a “speech community”, and to identify what sorts of divisions within that speech community can be observed in responses to a written questionnaire.

1.1 Goal 1: Document Features of Kansas City English

Our first research goal is rooted in traditional dialectology (or, as re-framed in Chambers (1994), “dialect topography”). There is a long precedent in language study for documenting speakers’ conscious knowledge of their language, either through a survey administered by a fieldworker or through a questionnaire completed by informants (see Chambers & Trudgill (1998: 15-21); Milroy & Gordon (2003: 51-56); or Dollinger (2015: 21-51) for detailed histories). While the written questionnaire suffered a period of disuse among linguists in the late-twentieth century, since the 1990s it has re-emerged as a viable tool for rapidly capturing large amounts of language data, especially following Chambers’s (1998) defense of questionnaires (also see Dollinger (2015: 53-86)).

We are particularly drawn to the variety of language features that the MLS allows us to report on. Previous studies of English in Kansas City—especially Lusk (1976) and Strelluf (2016, 2018)—examine phonetics and phonology to the exclusion of lexical and grammatical data. But surveying different language features may derive different dialectological conclusions. Labov, Ash & Boberg (2006) rely on phonological data to classify Kansas City as part of the Midland dialect region, but lexical data was the basis of Carver’s (1987) rejection of the Midland as a distinct dialect region, and Murray & Simon (2006: 15-28) offer a list of seventeen grammatical features that they argue define and validate the Midland’s regional status. With the exception of Ash’s (2006) report on responses from five Kansas Citians surveyed for the *Atlas of North American English* on a few lexical and grammatical items, we are unfamiliar with any published study of English in Kansas City beyond the realm of phonology.

MLS data, therefore, allow us to offer a much broader characterization of Kansas City English. While our treatment will be necessarily superficial given the number of items we hope to cover from the survey and the limits to the depth of information that can be gleaned from a written questionnaire, the presentation of this data will fill in important details about language in the community.

1.2 Goal 2: Explore the Role of Borders in the Community's Language Practices

Shortridge (2012: 204) describes Kansas City as “a community split by rivers, a state line, and race”. These geographical, political, and social borders that cut through the community affect Kansas Citians’ physical and psychological spaces in complicated ways. For instance, historically Kansas City was sited on the south bank of the Missouri River, just east of the bend where the river’s course turns from north-south to west-east. The bend came to mark the territorial border between Missouri and Kansas, so that to the north Missouri and Kansas were separated by the river, but to the south the states were separated by just a cartographic line. In Missouri, the river bisects Kansas City, and the “Northland” is the salient label for the entire region of Kansas City north of the Missouri River on the Missouri side of the border. Communities in Kansas, however, are not separated from each other by the Missouri River (the “Northland” label does not apply to any communities in Kansas, even if they are geographically northern). The communities in Kansas are also not separated from the communities in Missouri that are south of the river. Indeed, the only indication of the Missouri-Kansas border in the south is the mixed residential-commercial street, State Line Road.

While the communities in Kansas City south of the Missouri River are not physically separated, though, the Missouri-Kansas border nevertheless forms a significant boundary. The last quarter of the twentieth century saw a shift of population and economic prosperity from Missouri to Kansas (Shortridge 2012; US Census 2013). This shift has led to different popular valuations of Kansas City on either side of the state, with the Missouri side often being associated with poverty, danger, and bad schools and the Kansas side associated with wealth, safety, and good schools. Nicole_P_1972, who was interviewed for the projects reported in Strelluf (2016, 2018), had just moved from Kansas City, Missouri (MO) to a house just across the border in Kansas. She described the difficulty in moving from Missouri to Kansas in terms of class solidarity: “Unless you grow up here, it’s hard to understand how being a poor Missouri kid from a public school... like, how you have a chip on your shoulder”. She concluded that the move “wasn’t by choice” but was necessary for her family’s well-being.

State-border ideologies are codified locally in popular media, such as a “Kansas

City Barbie” meme that periodically circulates online in the area (we cite a 2008 posting, but found a number of more recent postings, and the first author received the meme by email in the early 2000s). In the meme, Barbies are stylized against Kansas City communities south of the Missouri River. Barbies from the Kansas communities of Leawood, Mission Hills, and Overland Park are connected to materialism, vanity, and nuclear family. E.g., Mission Hills Barbie “comes with an assortment of Kate Spade handbags, Lexus SUV, long-haired foreign dog named Honey and cookie-cutter house. Available with or without tummy tuck and facelift. Workaholic Ken sold only in conjunction with the augmented version”. Barbies from the Missouri communities of Belton, Grandview, and Independence are connected to crime, poverty, and dysfunction—e.g., Independence Barbie “comes with 9mm handgun, Ray Lewis knife, Chevy with dark-tinted windows, and Meth Lab Kit”. (Valuations of race—to be discussed below—also factor heavily in this meme.)

The Missouri-Kansas separation is reified by governmental actions. In recent years, Kansas City has gained notoriety as the site of a “border war”, where Missouri’s and Kansas’s competing tax incentives aimed at attracting jobs have made it possible for Kansas City companies to save huge sums of money by shifting office locations. By one estimate, over five years the tax incentives resulted in 10,000 jobs shuffling around Kansas City without any net job creation and a half billion dollars in lost tax revenue for Missouri and Kansas (“The New border war” 2014; “Jobs tug of war” 2016).

Racial divisions, especially between whites and African Americans, create yet another internal border. Across most of the twentieth century, combinations of racist real estate development (Brown & Dorsett 1978: 170-175; Worley 1990), school district boundaries (O’Higgins 2014), and white flight away from the urban core have deeply entrenched racial segregation in Kansas City. Troost, a north-south street in Kansas City, MO, is codified as the racial border, with African Americans living to the east and whites living to the west. A locally popular YouTube video, “A Tour of Kansas City”, illustrates the social salience of Troost in the psyche of white Kansas Citians (BottomsUpComedy 2013). At the 40-second mark, the video—which is otherwise a series of shots taken standing in front of local landmarks—is shot from inside a moving car. The “tour guide” says “This is Troost”. The “visitor” asks: “We going to stop?”. The tour guide shakes his

head, looks scared, and says only: “No”. The initial comments on the video refer to the hilarity of this joke, and to the need to mention streets farther east of Troost that also run through (perceptually) African American neighborhoods.

While Troost is perhaps the most symbolic racial border in Kansas City, the racist history it embodies has created parallel borders throughout the community. For instance, today’s North Kansas City, MO—which seeded the growth of the Northland—was initially populated in 1912 as a company town for the Armour and Swift meatpacking companies. In order to “attract ‘better’ workers”, the company banned African American, Mexican, and South European families from the city (Shortridge 2012: 75-76). While today’s demographic facts are, of course, not absolutely determined by historical foundations, the doctrine of first effective settlement (e.g., Zelinsky 1992) suggests that early strictures on populations may have outsized effects on future generations. In the 2010 Census, Clay County—the county that formed around Armour and Swift’s company town—had an African American population of 5.2 percent and LatinX population of 5.9 percent, compared with 23.9 percent and 8.4 percent in Jackson County, where Kansas City’s urban core is located (US Census 2016).

Finally, beyond these physical, political, and racial boundaries, psychological spaces seem to operate in nuanced ways. Strelluf (2018: 12) describes a qualitative change in attitudes toward Kansas City between interviews conducted during fieldwork in 2012-2013 and fieldwork in 2016. The former period was characterized by derisive comments like Seth_P_1972’s comparison of Kansas City to the idea of New York embodied in Frank Sinatra’s “New York, New York”: “If you can make it in Kansas City, you can just make it in Kansas City. [...] There’s no real prize”. The later interviews were characterized by Stephanie_1982’s description of herself as “completely Kansas City crazy” who believes “that we just live in awesome city and I never wanna leave”. Strelluf (2018: 12) speculated that these attitudinal changes might be linked to downtown revitalization projects and the Kansas City Royals winning the 2015 Major League Baseball World Series.

Strelluf (2018: 7-8) also qualitatively notes a generational change in attitudes toward Kansas City. He describes two relatively older interviewees, Carol_1959 and Molly_1973, who spoke of growing up in the Northland and pointedly denied having

lived in Kansas City, MO—though it turned out that Molly_1973, who described the Northland as “God’s Country”, lived in a house that actually was in Kansas City, MO. Carol_1959 and Molly_1973 are juxtaposed against two relatively younger interviewees, Danielle_1991 and Maya_1991, who both grew up in the Northland, but were both very particular about indicating that they were from Kansas City, MO and not the small independent enclave of North Kansas City, MO. These interviewees give the impression of orientations away from Kansas City, MO for the two older interviewees and toward Kansas City, MO for the two younger interviewees.

Similarly, the second author who moved to Kansas City from Wisconsin, impressionistically notes sharp divisions among friends and colleagues based on where they live relative to the Missouri River. People who live north of the river view the city south of the river variably as a “ghetto” or as a destination for special events like professional sports. People who live south of the river think of the area to the north as farmland that they drive through to get to the airport.

While these descriptions of perceptual borders in Kansas City are clearly unscientific, they suggest that Kansas Citians may be carving out spaces for themselves based on a range of internal and external factors. We want to study whether either established or perceived geographical, political, and social borders correlate with borders in linguistic practices.

1.3 Goal 3: Explore the Cohesiveness of the Speech Community

Our third goal is, in a sense, closely connected with Goal 2. In another sense, though, Goal 3 suggests a much more ambitious question for our research—and, regrettably, one that our data does not ultimately equip us to achieve. Nevertheless, we are motivated to look at MLS data in Kansas City as a way to explore a larger theoretical question of what constitutes a “speech community”.

Bucholtz (1999: 207) describes the “speech community” as “a central analytic tool of sociolinguistics” (see also Holmes & Meyerhoff 1999; Meyerhoff 2006). Referring to projects like Labov’s (1966/2006) foundational study of New York City English, she notes that through the lens of the speech community sociolinguists have successfully

identified quantifiable community-level patterns in language variation and change despite differences in linguistic practices at the individual level. Bucholtz advocates the “community of practice” as an alternative analytic unit, which examines micro-level linguistic practices, often emerging inductively among small groups of speakers who use language as part of a set of strategies to construct identities. (Bucholtz (1999) further credits Eckert & McConnell-Ginet (1992) for introducing the community of practice to linguistics.)

Eckert (2012) also contemplates the unit of variationist sociolinguistic analysis in her historicization of the field into three “waves”. She describes a first wave of large-scale, mostly urban studies that drew on sociological survey methods to treat speakers as “bundles of demographic categories” (Eckert 2012: 88) who engage in linguistic practices mostly passively in response to sociocultural norms. This was followed by a second wave that drew on network theory and ethnography, and that imbued speakers with agency for using language (among other) practices to navigate social interaction so language variation and change occur as a result of membership in locally defined social categories and movements of individuals among social groups. Finally, a third-wave recognized individuals as constructors of their own identities, and sociolinguists—often by working closely with individual speakers in a range of roles and interactions—identified language being operationalized as one of many stylistic factors to help people position themselves in their worlds.

Eckert (2012) does not use Bucholtz’s (1999) term “speech community”, but it is easy to see the speech community operating in Eckert’s historicization as something that shifts across the evolution of variationist sociolinguistics, from a geopolitical unit (e.g., the city), to a locally defined entity within a geopolitical unit (e.g., a school), to a group of friends, acquaintances, or colleagues (e.g., a circle of friends in a school in a city) within a locally defined entity. Likewise, the “community of practice” might be understood to exist as a continuum across the three waves. While a community of practice can be a very small group (as in the “nerd girls” of Bucholtz (1999) or a fraternity in Kiesling (1998), which Eckert (2012) cites as a third-wave study), a much larger group of people can also operationalize linguistic resources to engage in identity construction—as when a language variety becomes “enregistered” in the sense of Agha

(2003) (and as exemplified in Pittsburgh in Johnstone (2009), which Eckert (2012) also discusses in the context of third-wave variationist study).

Bucholtz's (1999) and Eckert's (2012) influential articulations of evolutions in sociolinguistics, then, suggest cyclically referential relationships among sociolinguists' methodological and theoretical commitments and their views of "communities" within these commitments. Studies like Podesva (2011) demonstrate this cyclical referentiality. Podesva (2011) examines one speaker's operationalization of features of the California Vowel Shift (CVS) (e.g., Eckert 2004; Kennedy & Grama 2012; Fridland *et al.* 2016) to index different identities in interactions with a group of friends, a single friend, and a professional supervisor. The establishment of the baseline features of the CVS that the speaker makes use of is firmly rooted in first-wave variationist approaches (in Eckert's (2012) terminology), and CVS features exist at the macro-level of the San Francisco English speech community (in Bucholtz's (1999) terminology). However, the focus on a single speaker's construction of meaning and identity through these features is clearly embedded in third-wave variationist approaches and relies on understanding the speaker as moving among different communities of practice.

In one sense, then, variationists working in different "waves" or in different levels of community can (and certainly do) provide complementary perspectives. First-wave variationists might describe the features that are present in a city, second-wavers might identify how those features advance and take on meaning in a school in the city, and third-wavers might discover what a clique in the school does with the features. This gives rise to circularity, though. If individuals make up communities and communities provide the linguistic backdrop for individuals, how can we know when we are identifying individual practices or when we are identifying community practices? Put another way, when are we dealing with the speech community, and when are we dealing with a community of practice?

We are not, of course, making a new observation about the field. Indeed, we seem to have presented a less articulate restatement of what Labov (1972: 185-187) referred to as the Saussurian Paradox. The question of what constitutes a speech community carries tremendous consequences for decisions researchers make as they design studies and draw conclusions from them. Can a researcher—short of decades-long projects

involving scores of researchers like those that Labov and colleagues have conducted in Philadelphia since the early 1970s—make meaningful claims about language practices in a city? Without such a backdrop of community language norms, can researchers make claims about group or individual linguistic practices or performances?

We will make no claims to being able to answer such high-level theoretical questions with questionnaire data. Nevertheless, as we probe the MLS for data about language in Kansas City and for data about language practices across borders in Kansas City, we are conscious of a larger question. On an important level, we are asking to what extent can we think of Kansas City as a speech community?

2. Methods

Our survey is included as Appendix A. Surveys were administered by students enrolled in the first author composition and linguistics courses at Northwest Missouri State University during 2015 and 2016. Students were assigned to recruit adult informants who grew up in Missouri or within 50 miles of the Missouri border. Informants provided data on where they had lived (both in and out of Missouri), age, sex, and ethnicity. We did not collect data in the survey to tie informants to socioeconomic categories like occupation. We report from 477 surveys where informants listed as their hometown a community within the Kansas City Metropolitan Statistical Area (US Census 2015).

The survey consisted of three sections that solicited judgments about the phonological statuses of vowels, lexical items, and grammaticality. Items in each section were chosen for a variety of reasons, including to test against previous studies of Missouri speech (e.g., Faries 1954; Faries 1967; Faries & Lance 1993; Lance & Faries 1997; Strelluf 2016, 2018), to test against national dialect surveys (e.g., Vaux & Golder 2003), and to empirically check researchers' casual observations of Missouri speech.

Several survey items elicited a wide range of responses from informants and required some re-coding to facilitate statistical analysis.

2.1 Ethnic Background

Besides anticipated responses like “white”, “African American”, and “Hispanic”, the survey item also generated indefinite labels like “unknown”, “none”, and “American”, as well as labels that likely indicated family historical narratives rather than informants’ modern participation in ethnic communities (e.g., “white, Scottish, Native American”). Responses for ethnicity were re-coded to an “interpreted race” column, that categorized informants as African American, Asian, biracial, LatinX, Native American, and white. Henceforth we use the label “race” to refer to this re-coded category.

2.2 Hometown

To provide a way to group and compare communities, columns were added to capture which side of the Missouri-Kansas border a city is located on, whether the city is north or south of the Missouri River, and what county a city is located in. Each of these seemingly obvious classifications presented some challenges.

The category of “county” was complicated by communities that spread across multiple counties. Six relatively small cities in the survey—De Soto in Kansas and Excelsior Springs, Holt, Lawson, Oak Grove, and Smithville in Missouri—contain large rural tracts that account for spread across county boundaries. These six cities were coded for county where most of their populations and government services were located. Kansas City, MO presented a different case. The city was originally seated in Jackson County, but annexed areas in Cass, Clay, and Platte Counties between 1950 and 1963 (Shortridge 2012: 103). “KC” was, therefore, included as a category for county to account for responses of “Kansas City” as a hometown.

Responses of Kansas City as hometown also created problems for classifying “state” and “river”. Since informants may use labels like “KC” or “Kansas City” to indicate the Kansas City MSA regionally as their hometown, these labels provide no information on where informants actually come from within the Kansas City area. Responses of “KCMO”, “Kansas City, MO” and “KCK” were classified according to state

as Missouri or Kansas, but state was coded as “unknown” for labels of “KC” and “Kansas City”. “KCK” was coded as south of the Missouri River, but all other labels for Kansas City were left as “unknown” for the variable of river. As such, the label “Kansas City” must be thought of in this study as a place where informants identify being from, rather than as a true indication of the city informants grew up in.

2.3 Please list any other places you have lived and indicate when in your life you lived in each place

Responses were converted to a binary “yes” or “no” corresponding to whether an informant had provided a response. We use the label “other” for this category.

2.4 Lexical Items 8, 10, and 13

Each of these survey items elicited not only a wide range of responses, but also a wide variety of spellings for the same underlying lexemes (e.g., “Boot heal” for *Bootheel*, “teeder tawder” for *teeter-totter*). While divergent spellings offer dialectological and sociolinguistic insights, for the sake of statistical analysis, spellings were standardized. We also describe some additional ad hoc simplifications of responses to survey items below.

We detail these coding decisions because they will necessarily affect results we generate. We re-coded with great concern for representing responses in ways that were consistent across the sample and with social and linguistic realities. We were particularly concerned about choices for re-coding ethnicity. While it is appropriate to eliminate non-meaningful differences in a dataset, it is impossible to know *a priori* which differences are meaningful. The imposition of order on a sample homogenizes responses in ways that might obscure, eliminate, or even create sociolinguistic facts—or, worse, dehumanize the people who provided those responses. We report data from this sample noting that we made decisions to re-code carefully and critically, but also acknowledge that our coding decisions shaped the data.

Table 1 lists counts for each demographic category in the final dataset. It does not

include the 40 unique cities listed by informants as “hometown”.

| Sex | | River | | State | | Other | |
|--------|-----|---------|-----|----------|-----|-------|-----|
| Female | 278 | North | 104 | Kansas | 17 | No | 278 |
| Male | 199 | South | 194 | Missouri | 381 | Yes | 199 |
| | | Unknown | 179 | Unknown | 79 | | |

| Age | | Race | | County | |
|-------|-----|------------------|-----|-----------|-----|
| 18-19 | 172 | African American | 54 | Bates | 3 |
| 20-25 | 120 | Asian | 8 | Cass | 36 |
| 26-40 | 41 | Biracial | 11 | Clay | 68 |
| 41-55 | 53 | LatinX | 24 | Clinton | 4 |
| 56-85 | 26 | Native American | 5 | Jackson | 134 |
| | | White | 375 | Johnson | 9 |
| | | | | KC | 179 |
| | | | | Lafayette | 12 |
| | | | | Platte | 31 |
| | | | | Ray | 1 |

Table 1. Demographic characteristics of informants

Our sample is skewed in several categories, reflecting our reliance on college students to collect surveys. Nearly two-thirds of informants are 25 years old or younger, which is unsurprising since the students collecting surveys were mostly in this age range, and likely often turned to peers for help with the assignment. We have a higher proportion of informants identifying as female than male. Missouri is much better represented than Kansas—likely a result of Missouri students receiving in-state tuition rates at the university, so that the student population contains many more Missourians than Kansans. In counties, the mostly urbanized counties like Jackson and Clay are much more strongly represented than lightly populated rural counties like Bates and Ray. However, Johnson County, which is in Kansas, is mostly urbanized and is under-represented in our sample.

The sample is skewed toward white informants, reflecting the racial homogeneity of Northwest Missouri State University. This racial skew, however, is not terribly far off

the demographics of Kansas City. Census estimates for 2014 of the counties from which we report data are 74.1 percent white, 13.4 percent African American, 8.8 percent LatinX, 0.5 percent Native American, and 3.3 Asian (MARC 2017). Our survey sample is 78.6 percent white, 11.3 percent African American, 5 percent LatinX, 1 percent Native American, and 1.7 percent Asian.

We analyzed the sample in R (R Core Team 2016) through conditional inference trees (first described in Strasser & Weber 1999). We worked from the “party” package (Hothorn, Hornik, Strobl & Zeileis 2017) and its Conditional Inference Trees (ctree) (Hothorn, Hornik & Zeileis 2006) and Conditional Random Forests (cforest) (Hothorn, Buehlmann, Dudoit, Molinaro & Van Der Laan 2006; Strobl, Boulesteix, Zeileis & Hothorn 2007; Strobl, Boulesteix, Kneib, Augustin & Zeileis 2008) functions. Conditional inference trees mine responses to a dependent variable for statistically significant predictors. When significant predictors are found, responses are split at the factor with the lowest p -value, forming a node with two branches of data. Each branch is then examined for significant predictors. The process repeats recursively until it exhausts all significant splits. Thus, it forms a hierarchical explanation for responses to a variable.

Random forests provide a check against overfitting data in conditional inference trees by growing and comparing multiple trees. Each tree is built from a bootstrapped sample of the dataset and, as the tree is constructed, only a random subset of variables is made available to test for splits at each node. In this manner, the algorithm builds many different trees based on slightly different datasets. The function then compares all trees to determine factors that hold the most explanatory weight for responses to the dependent variable. Finally, it tests a subset of the data against these factors to validate their power in predicting responses, resulting in an “importance measurement” for each factor. Stephens (2014) provides an accessible explanation and tutorial on random forests. Conditional inference trees and random forests have been used in a few sociolinguistic studies (e.g., Chevalier 2016; Gordon *et al.* 2004; Tagliamonte & Baayen 2012).

Each item in the MLS was modeled as a conditional inference tree with the independent variables of age, sex, race, river, county, state, and other. Dependent variables that showed significant interactions were modeled as random forests, with

2000 trees being generated for each model and three variables tested at each node. Importance measures were derived with the Variable Importance `varimp()` subfunction of `cforest()`. Each random forest model was also rebuilt with a second “random seed” (per `cforest()` documentation) to check the validity of the first set of importance weights. There were no differences in rankings returned by these validation runs. Random forests were also built with the variable `city` replacing `county`, `river`, and `state`. However, models with `city` never resulted in substantial differences from models with `county`, `river`, and `state`, so we do not discuss them.

Random forests provide a sanity check on individual conditional inference trees, allowing us to make more confident claims about the importance of social variables in predicting survey responses. In a few cases a conditional inference tree returned a significant predictor for a variable, but random forest models selected other predictors as important. We presume these to be cases of the conditional inference tree overfitting data and treat these as variables without significant predictors.

As we present data below, we initially refer to the importance measures returned by random forest models to establish an overall context for variables. We then return to conditional inference trees to highlight results for specific survey items because they offer intuitive models of data.

3. Results

Importance measures from a random forest model for each variable are printed in Tables 3, 4 and 6. All are “seeded” with the same random seed. The specific importance measures should be thought of as ways to compare the relative weights of social factors in predicting responses to each survey item. They should not be treated as absolute weights because a different random seed will result in slight differences in importance measures (though, as noted above, we checked for fluctuations in rankings for each variable as a result of random seeds, and no such fluctuations occurred).

To our knowledge there is not an established value that marks an importance measure as “important”. In test runs of models with seven dummy predictor variables,

one of which was perfectly correlated with a dependent variable and six of which were uncorrelated, the correlated predictor returned importance weights around 0.15 and the uncorrelated predictors returned weights of 0. When the model was run with just the dependent variable and the perfectly correlated variable, it returned an importance weight around 0.67. So, bigger numbers show increased predictive weight, but there is not a threshold for a factor being “important”.

3.1 Controls

The MLS included five items meant to test the effects of the survey instrument itself on responses. The minimal pairs *wait-weight*, *caught-coat*, and *here-hear* showed no significant interactions for social factors. The grammatically standard negative-polarity *anymore* in “Kids don’t get enough exercise *anymore*” also showed no interactions. The minimal pair *whole-hole* showed very small importance measures for race and other (see Table 2). The conditional inference tree shows a statistically significant break for race ($p < 0.001$), with the small sample of Asian Americans being slightly more likely than all others to claim the pair sound “different” and less likely than all others to claim they sound “close”. The small quantitative difference and low importance measures, though, show that this split for race is quite marginal.

Gordon (2006) used such controls to examine the reliability of responses to a written survey about the phonological status of vowels. As with Gordon (2006), informants in the data reported here answered controls “incorrectly”, indicating differences between sounds that would be expected to be homophonous for American English speakers, similarities for traditionally distinct *caught* and *coat*, and denials of having heard or used an uncontroversially standard grammatical construction. In fact, Gordon (2006: 63) reports that 19 percent of respondents to his survey marked the pair *whole-hole* “distinct”. In the present study, 20.8 percent gave a response other than “same” for the pair, suggesting a very similar level of sensitivity to the written survey instrument between studies. Crucially, though, except for a relatively minor observed difference according to race for *whole-hole*, the data reported in this study is homogenous in the ways that informants answered these controls. In other words,

across the survey, everybody got the same wrong answers. This indicates uniformity across demographic factors in approaching the questionnaire.

3.2 Phonological Items

Table 2 shows responses to minimal pairs items. Informants were asked to indicate whether each pair sounded the “same”, “close”, “different” or “either”. (The option for “either the same or distinct” was included to test a hypothesis that will be reported elsewhere. We will not discuss it here but will presume that a response that vowels can sound “either the same or distinct” is suggestive of a loss of phonological distinctness). Table 3 provides importance measures for each minimal pair. Importance measures are quite small, indicating a fair degree of uniformity in responses.

| | Different | Close | Same | Either | NO RESPONSE |
|--------------|-----------|-------|------|--------|-------------|
| wait-weight | 15 | 68 | 389 | 4 | 2 |
| haul-hull | 339 | 110 | 26 | 2 | 0 |
| caught-coat | 426 | 29 | 9 | 10 | 3 |
| cot-caught | 63 | 85 | 323 | 5 | 1 |
| gym-gem | 144 | 130 | 194 | 8 | 1 |
| whole-hole | 23 | 66 | 376 | 10 | 2 |
| pull-pool | 289 | 114 | 62 | 9 | 3 |
| dawn-Don | 46 | 103 | 319 | 9 | 0 |
| pen-pin | 128 | 132 | 201 | 13 | 3 |
| Polly-Paulie | 68 | 108 | 286 | 10 | 5 |
| goal-gull | 256 | 162 | 52 | 3 | 4 |
| here-hear | 24 | 56 | 386 | 7 | 4 |
| tour-tore | 283 | 106 | 78 | 6 | 4 |
| bull-bowl | 274 | 120 | 72 | 9 | 2 |

Table 2. Responses to minimal pairs (controls shaded)

| | Age | Sex | Race | River | County | State | Other |
|--------------|---------------------------|--------|-------|--------|--------|--------|--------|
| wait-weight | no significant predictors | | | | | | |
| haul-hull | no significant predictors | | | | | | |
| caught-coat | no significant predictors | | | | | | |
| cot-caught | no significant predictors | | | | | | |
| gym-gem | no significant predictors | | | | | | |
| whole-hole | <0.001 | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | 0.002 |
| pull-pool | <0.001 | <0.001 | 0.019 | <0.001 | 0.001 | <0.001 | <0.001 |
| dawn-Don | 0.002 | 0.002 | 0.008 | <0.001 | 0.002 | <0.001 | <0.001 |
| pen-pin | 0.002 | 0.011 | 0.018 | 0.010 | 0.005 | <0.001 | <0.001 |
| Polly-Paulie | tree and forest mismatch | | | | | | |
| goal-gull | 0.015 | <0.001 | 0.005 | <0.001 | <0.001 | 0.010 | 0.008 |
| here-hear | no significant predictors | | | | | | |
| tour-tore | tree and forest mismatch | | | | | | |
| bull-bowl | no significant predictors | | | | | | |

Table 3. Importance measures for minimal pairs (controls shaded)

Six minimal pairs—*haul-hull*, *cot-caught*, *gym-gem*, *Polly-Paulie*, *tour-tore*, and *bull-bowl*—show no significant effects from social factors. “Different” is the majority response for the pre-/l/ conditional mergers examined with *haul-hull* and *bull-bowl*. A majority also select “different” for the pair *tour-tore*. “Same” is a strong majority response for the low-back pairs *cot-caught* and *Polly-Paulie*. Responses to the pre-nasal pair *gym-gem* are relatively evenly split among “different”, “close” and “same”.

Race is the largest importance factor for four minimal pairs: *pull-pool*, *dawn-Don*, and *pen-pin*. For the low-back vowel pair *dawn-Don*, African American and Native American informants are more likely than others to judge the pair “close” ($p < 0.001$). Among other informants, the age of 59 is also a significant divide ($p = 0.038$), with

younger informants strongly indicating the pair sounds “same”. In the pre-nasal conditional merger of *pen-pin*, Asian American informants show a stronger preference for “different” ($p < 0.001$), as do LatinX and Native American informants ($p = 0.019$), while all other informants prefer “same”. Age predicts responses to *goal-gull*. Informants older than 57 judge the pair “different”, while younger informants increasingly judge the pair “close” and, to a lesser extent, “same” ($p < 0.001$).

Figure 2 shows the conditional inference tree for the minimal pair *pull-pool* to explore one survey item. Surveys that left this item blank are dropped from the model. It shows a significant split at race, with informants identified as African American, biracial, and LatinX showing a higher proportion of responses of “close” and “same” than other informants. Node 5 splits again along racial lines, with the 11 biracial informants showing a strong preference for judgments of “close” and with African Americans and LatinX informants showing a relatively high proportion of “same” responses. In the left branch of the tree at Node 2, the Asian, Native American, and white informants split according to age. Informants older than 31 show a strong preference for judgments of “different”, while younger informants show an increase in judgments of “close” and “same”.

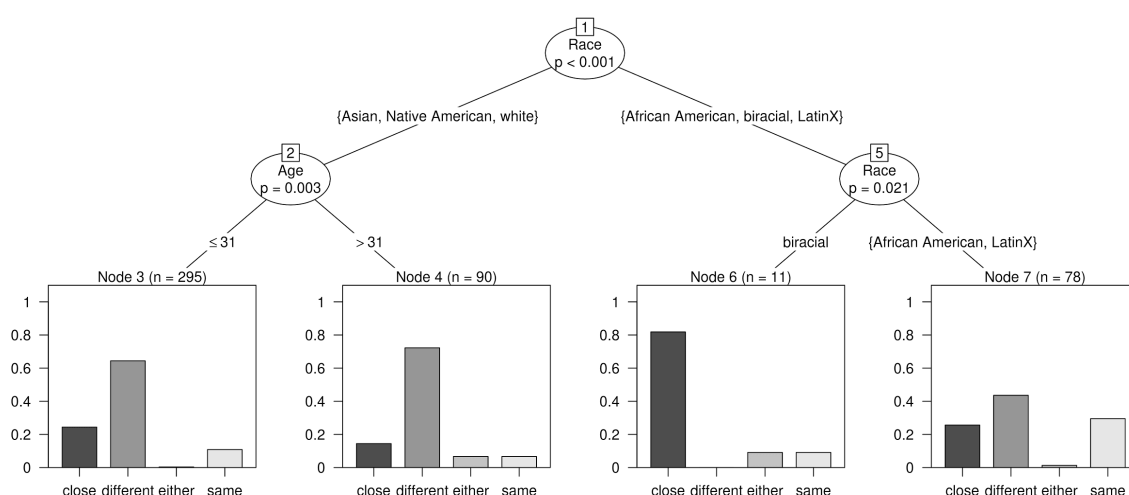


Figure 2. Conditional inference tree for *pull-pool*

3.3 Lexical Items

This section reports on survey items that examine both traditional lexical variables (i.e., what word does someone use for something) and lexeme-specific phonological variables (i.e., how does someone pronounce a specific word). The importance measures for both types of lexical item are printed in Table 4. In an initial run, all nine lexical items surveyed appear to result in significant effects from social factors. However, the tiny importance measures for age and race in the pronunciation of *crayon* result in a significant split that separates just the small sample of Asian Americans from other Kansas Citians. (Disyllabic *crayon* that rhymes with “play on” is the strongly dominant form, followed by monosyllabic “brown” in a distant second.) For the term for “someone who thinks he’s better than everyone else”, the importance measure for age is a result of the 72 unique lexemes elicited by the open-ended question on the survey. When the model is simplified to the dominant terms (*arrogant*, *cocky*, *conceited*, *egotistical*, *full of themselves*, *stuck up*) and “other”, the effect disappears. As such, these two survey items should be treated as, at best, marginally significant.

| | Age | Sex | Race | River | County | State | Other |
|---------------------------|--------|--------|--------|--------|--------|--------|--------|
| coupon | 0.002 | 0.026 | <0.001 | <0.001 | 0.002 | <0.001 | <0.001 |
| crayon | 0.002 | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | <0.001 |
| grocery | 0.014 | <0.001 | 0.010 | <0.001 | <0.001 | 0.001 | 0.002 |
| playground equipment | <0.001 | <0.001 | 0.053 | 0.006 | 0.002 | <0.001 | 0.002 |
| better than everyone else | 0.008 | 0.003 | <0.001 | <0.001 | 0.004 | <0.001 | <0.001 |
| southeast Missouri | 0.001 | 0.013 | 0.019 | 0.010 | <0.001 | <0.001 | <0.001 |
| highway | 0.009 | <0.001 | <0.001 | 0.020 | 0.004 | 0.024 | 0.003 |
| anyway(s) | 0.037 | <0.001 | 0.003 | 0.006 | 0.001 | 0.009 | 0.002 |
| lobster | 0.011 | 0.008 | 0.037 | 0.012 | 0.002 | <0.001 | <0.001 |

Table 4. Importance weights for lexical items

Importance measures suggest a greater variety of social influences on responses to lexical items than was observed for minimal pairs. Race is counted as the largest importance measure for four items, age is largest for two, and sex and state are largest for one each.

Importance measures for the survey items are also relatively large when compared with importance measures for minimal pairs. The effect of race on the name for “playground equipment consisting of a long board with seats on the ends” is strikingly large. The open-ended question elicited six lexemes (*balance*, *levy*, *seesaw*, *teeter board*, *teeter-totter*, *totter*), several presumably misunderstood responses of *slide* and *swing*, and a few blank responses. A simplified version of this survey item that reduces responses to *seesaw*, *teeter-totter*, “both” and “other” is modeled as a conditional inference tree in Figure 3. It shows that *teeter-totter* is the preferred form among informants identified as Asian American or white, while *seesaw* (and various other forms) are more prevalent among other informants. *Seesaw* is the more dominant form especially among African American and biracial informants.

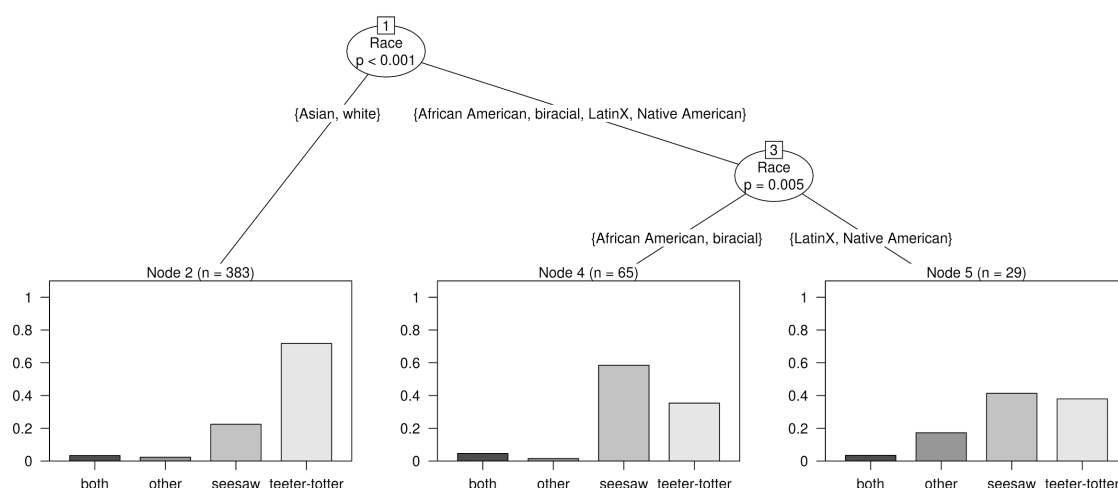


Figure 3. Conditional inference tree for the name for playground equipment

The name for a “small lobster-like animal that can be found in freshwater creeks and lakes” generated an even larger number of unique terms at 31, including the expected forms like *crawdad*, *crawfish*, and *crayfish*, as well as a number of responses showing unfamiliarity with the animal (e.g., *crab*, *shrimp*, and variations of “I don’t

know”). (There were also terms that showed definite awareness of the animal such as *poor man’s lobster*.) A simplified scale that reduces responses to *crawdad*, *crawfish*, *crayfish*, “IDK” (for “I don’t know”), and “other” splits at race ($p < 0.001$), with LatinX and white informants favoring *crawdad*. The responses of African American, Asian American, biracial, and Native American responses are further split by sex, with *crawfish* as a preferred form among women and *crayfish* and “other” being preferred among men.

The term for the “extreme southeast corner of the state of Missouri” was also solicited by open-ended question and resulted in 41 unique labels. Some of these provide very obvious social commentary on attitudes toward that region of the state, including *armpit*, *Bible Belt*, *hicksville* and *Missoura*. While interesting, these overtly evaluative labels are beyond the purview of this study. When responses are simplified to the traditional labels *Bootheel* and *Boot Hill*, as well as the apparent compromise *Boot*, “IDK” and “other”, a significant split opens between whites and Native Americans versus all others ($p < 0.001$). “IDK” is the leading response among the non-white/Native American group. The white and Native American node is further split at river ($p = 0.025$). *Bootheel* is the dominant term for both these groups, but the “unknown” category of informants who listed Kansas City as their hometown have a relatively high proportion of “IDK”, while the informants who could be placed either north or south of the Missouri River have a relatively higher proportion of *Boot Hill*. In other words, for people who claim Kansas City as their hometown and for non-white Kansas Citians, the southeast corner of Missouri is often unnamed territory.

The pronunciation of the vowel in first syllable of *coupon* shows a strong effect for sex ($p = 0.02$), with a glide onset /ju/ being the majority response among females. Among males, those 57 years or younger have a higher proportion of /u/ without a glide (i.e., yod-dropping), while older males retain the glide ($p = 0.01$). The pronunciation of the orthographic <c> in *grocery* splits for age at 43 ($p = 0.017$), with younger informants showing a strong preference for /j/, while older informants are more nearly split between /j/ and /s/. The choice between *anyway* and *anyways* is also strongly affected by age, with a cutoff at 26 ($p < 0.001$). Older informants show a strong preference for *anyway*; younger informants are split almost evenly between *anyway* and *anyways*.

Finally, in Kansas City numbered highways are often referred to with the number

preceding *highway*, rather than with *highway* preceding the number as is standard elsewhere in the United States. Thus, *Highway 24* might be called *24 Highway* in Kansas City. The conditional inference tree plotted in Figure 4 shows this particular lexical variable is differentiated by the Missouri River. The “Kansas City” production of *24 Highway* is more accurately described as a feature of Kansas City English south of the Missouri River, while *24 Highway* and *Highway 24* are given as responses in nearly equal proportions north of the river and among informants who list anomalous Kansas City as their hometown ($p=0.002$).

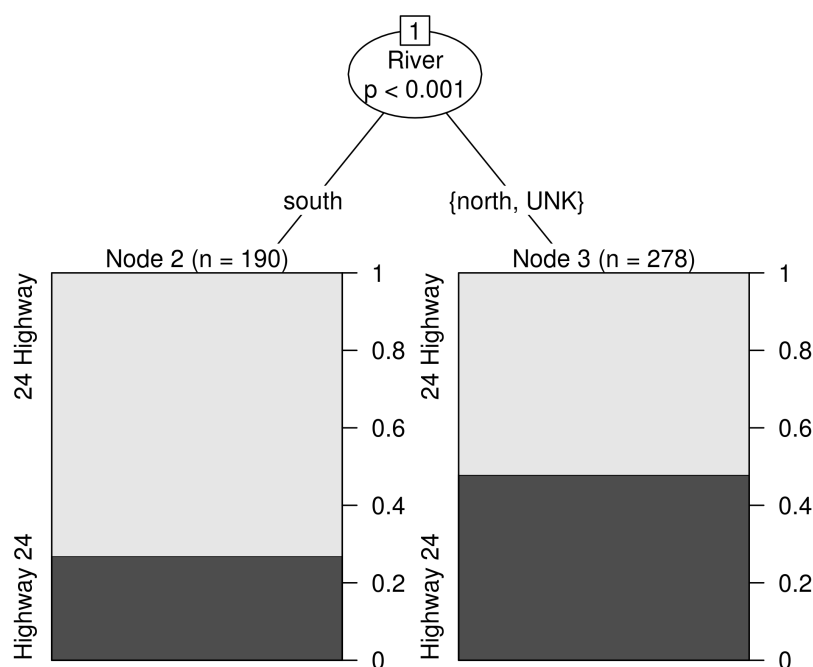


Figure 4. Conditional inference tree for *Highway 24* vs. *24 Highway*

3.4 Grammaticality Items

Table 5 lists responses to eight grammatical items. Informants were asked to indicate whether each sounded like a “phrasing that you might use in everyday speech; phrasing that you have heard from other people but that you wouldn’t use yourself; or phrasing that you’ve never heard before”. Table 6 shows the grammatical variables’ importance measures.

| | Use | Heard | Never | NO RESPONSE |
|--------------------------------------|-----|-------|-------|-------------|
| <i>come with</i> ∅ | 275 | 179 | 23 | 0 |
| punc. <i>whenever</i> , broad frame | 109 | 301 | 65 | 2 |
| pos- <i>anymore</i> , clause-final | 93 | 213 | 170 | 1 |
| neg- <i>anymore</i> | 234 | 214 | 29 | 0 |
| punc. <i>whenever</i> , narrow frame | 159 | 221 | 97 | 0 |
| pos- <i>anymore</i> , clause-initial | 79 | 183 | 212 | 3 |
| <i>needs</i> +PAST | 264 | 157 | 56 | 0 |
| <i>want</i> +PREP ADV | 294 | 149 | 33 | 1 |

Table 5. Importance weights for grammaticality judgments (control shaded)

| | Age | Sex | Race | River | County | State | Other |
|--------------------------------------|---------------------------|--------|-------|--------|--------|--------|--------|
| <i>come with</i> ∅ | 0.106 | <0.001 | 0.010 | <0.001 | 0.006 | <0.001 | <0.001 |
| punc. <i>whenever</i> , broad frame | no significant predictors | | | | | | |
| pos- <i>anymore</i> , clause-final | 0.006 | <0.001 | 0.049 | <0.001 | <0.001 | <0.001 | 0.001 |
| neg- <i>anymore</i> | no significant predictors | | | | | | |
| punc. <i>whenever</i> , narrow frame | tree and forest mismatch | | | | | | |
| pos- <i>anymore</i> , clause-initial | 0.031 | 0.002 | 0.026 | <0.001 | <0.001 | <0.001 | 0.011 |
| <i>needs</i> +PAST | <0.001 | 0.007 | 0.020 | 0.015 | 0.002 | <0.001 | 0.007 |
| <i>want</i> +PREP ADV | <0.001 | 0.001 | 0.029 | 0.010 | 0.005 | 0.004 | 0.007 |

Table 6. Importance weights for grammaticality judgments (control shaded)

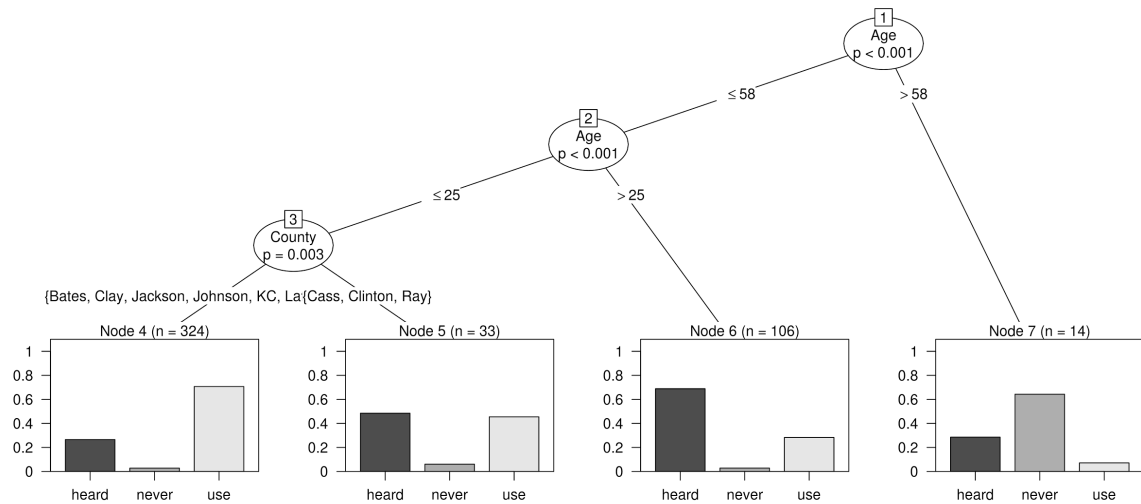
As noted above, the control *anymore* as a negative polarity item has no significant predictors. Punctual-*whenever*—the use of *whenever* to refer to a specific or known moment in time—also shows no significant predictors, with the majority of Kansas Citians indicating they have heard phrases like the two instances included in the MLS.

Two forms of positive-*anymore*—the use of *anymore* without a negative polarity

item to license it—differ in their interaction with social factors. In the item “food is so expensive *anymore*”, positive-*anymore* occupies clause-final position, and is strongly predicted by race. Informants interpreted as Native American or white have the highest proportion of “heard” responses ($p=0.005$). African American, Asian American, biracial, and LatinX informants are further split by state ($p=0.002$), with respondents from Kansas and “unknown” overwhelmingly indicating they have never heard the construction. The remaining Missourians are split again for race ($p=0.02$), biracial informants are divided between “heard” and “use” and others respond mostly with “never”. Clause-initial positive-*anymore* in “*Anymore* most things are made in China” also splits at race ($p=0.001$), with Asian American, Native American, and white informants reporting a greater proportion of “heard”. African American, biracial, and LatinX informants are split at age ($p=0.005$), with those over 23 reporting a plurality of “use”, while younger informants report a higher proportion of “never”.

The *needs*+PAST PARTICIPLE of “The car *needs* washed” shows a strong effect for race ($p<0.001$). African American and Asian American informants respond with “use”, “heard” and “never” in nearly equal proportions, while other racial groups including the white majority have a very high proportion of “use” responses and very low proportion of “never”. The construction of *want*+PREPOSITIONAL ADVERB in “I *want off* this bus” also shows an effect for race. A majority of African Americans respond with “heard”, while all other Kansas Citians show a very high proportion of “use”.

Finally, the stranded preposition construction of *come with* \emptyset in “We’re going out. Do you want to *come with*?” patterns as a change in apparent time. Figure 5 displays its conditional inference tree.

Figure 5. Conditional inference tree for *come with Ø*

Come with Ø shows two highly significant splits for age ($p < 0.001$). Informants older than 58 generally report “never” having heard the feature. Informants between 25 and 58 primarily report having “heard” *come with Ø* and a growing proportion report they would “use” it ($p < 0.001$). Informants under 25 are further divided by county ($p = 0.003$). The largely rural outlying counties of Cass, Clinton, and Ray have nearly equal proportions of “heard” and “use”, while “use” becomes the dominant response among other younger Kansas Citians, including those from the more urban counties of Jackson, Johnson, and Clay. These splits combine to show familiarity and acceptance of *come with Ø* increasing in apparent time, and potentially show the feature being more closely associated with urban speech than non-urban speech.

3.5 Revisiting Race

While most social factors showed some role in predicting responses, the variable of race emerged most often. To offer an alternative view of interactions between social factors and survey responses, we ran “flipped” versions of our model, entering each social factor in turn as a dependent variable and all survey items as predictor variables.

The results of this approach were quite striking. County, river, and state selected no items as significant. Sex selected “What do you call the small lobster-like animal that can be found in freshwater creeks and lakes?” as significant. Age selected *come with Ø*,

as well as clause-initial positive-*anymore*, *want off* NP, and *anyway(s)*. Race, by contrast, selected twelve significant splits according to survey items. Among them are the minimal pairs *gym-gem* and *haul-hull*, the pronunciation of *crayon*, the name for playground equipment, positive-*anymore*, and punctual-*whenever*. The details of the model are less important than the confirmation that model provides that race is the major predictor of differing responses to the MLS.

We therefore reran all of our models for the two major racial groups in our survey, whites and African Americans, to examine whether more nuanced linguistic practices would emerge in racially homogenous datasets. Among white informants, the minimal pairs *pull-pool*, *dawn-Don*, *goal-gull*, and *bull-bowl* all select age as significant. Each now emerges as a change in apparent time, with judgments of “close” and “same” displacing judgments of “different”. The pronunciation of the vowel in *coupon*, *anyway(s)*, and *come with* \emptyset model as described above for the full sample. Among African American informants, only the pronunciation of <c> in *grocery* has a significant predictor. It models as a change in apparent time, with /s/ holding a slight majority for informants older than 20 while /j/ is strongly preferred for informants 20 and younger.

While this reworking of the data offers some new insights into dialectology in Kansas City, it offers a strong confirmation of the role of race in predicting responses. Race is frequently selected as a significant predictor of survey responses in models, race itself selects many linguistic factors as predictors, and, when racial groups are isolated, survey responses become relatively homogeneous..

4. Discussion

We entered this project with three research goals. We discuss each of them in turn here in light of results from the MLS.

4.1. Goal 1: Document Features of Kansas City English

MLS data allowed us to examine—even if only superficially—a wide range of

phonological, lexical, and grammatical variables in Kansas City. Our findings in some cases confirm other studies of language in the area, and in other cases offer new insights into the dialect topography of Kansas City.

Minimal pairs items examined several sets of vowels for phonological mergers. These items offer us a number of bridges between the present work and previous research. Responses suggest the low-back pairs *cot-caught*, *dawn-Don*, and *Polly-Paulie* to be undergoing merger in informant judgments. Relative to Gordon (2006: 62), we report higher rates of judgments of “same” for *cot-caught* in Kansas City (41.3 percent of responses in Gordon versus 67.7 percent here) and *dawn-Don* (41.4 percent in Gordon versus 66.9 percent). On the other hand, we found a lower rate of judgments of “same” for *Polly-Pauley* than Strelluf (2018: 185-186: Appendix C.2) (75.5 percent versus 60.0 percent), as well as for *dawn-Don* (78 percent in Strelluf versus 66.9 percent here). We found similar rates of “same” judgments for *cot-caught* (70 percent in Strelluf versus 67.7 percent). Despite these differences, MLS responses lead us to the same conclusion as prior studies of judgments about the low-back merger in Kansas City: it is very advanced.

Strelluf (2016) examined productions of the vowels in the minimal pairs *haul-hull*, *pull-pool*, *goal-gull*, and *bull-bowl*, as well as speaker judgments of the minimal pairs *pull-pool* and *bull-bowl*. Strelluf (2016: 390) found that closer productions of *haul-hull* represented an older vowel configuration—an observation supported by the preponderance of “different” judgments reported here. Strelluf (2016: 388) found productive overlap of *goal-gull* was widespread; in MLS data, the 34.0 percent of informants who judge this pair “close” suggests that this productive overlap is not yet sufficient to collapse the conscious phonemic distinction between vowels in this environment, but that the distinction is weakening among younger Kansas Citians. Strelluf (2016: 395) also found that judgments of *pull-pool* and *bull-bowl* as “close” or “same” were becoming more common among younger Kansas Citians. The present research supports the conclusion generally that judgments of “close” are becoming more common in apparent time for these pairs and connects judgments of *pull-pool* as “close” especially with African Americans and judgments of *bull-bowl* as “close” with white Kansas Citians.

Strelluf (2018: 188-189: Appendix C.4) reports judgments for the minimal pairs *gym-gem* and *pen-pin*. A smaller proportion of MLS informants judged *gym-gem* “same” (56 percent in Strelluf versus 40.7 percent here), while the two studies show very similar levels of “same” judgments for *pin-pen* (44 percent in Strelluf versus 42.1 percent). Strelluf (2018: 106) speculates that social evaluations of this pre-nasal conditional merger might be blocking its progress among younger Kansas City women—particularly ways that the merger might index southernness or rurality. The present research offers no support for this idea, since these items do not pick out social factors as predictors of responses.

Finally, this study finds a strong preponderance of judgments of *tour-tore* as “different”. To the best of our knowledge, this pair was previously unstudied in Kansas City.

MLS responses show lexical items to be a rich area of differentiation within Kansas City. Our primary point of comparison here is the maps stored from the Harvard Dialect Study (Vaux & Golder 2003), to which we can now add a great deal of detail and nuance for one community.

We treated the pronunciation of the first vowel in *coupon* as a lexical variable, though it could be an entrypoint to a phonological study of yod-dropping (cf. Chambers 1998 or Dolinger 2015: 82-86). Vaux & Golder’s (2003) map shows both /ju/ and /u/ present in Kansas City. We have confirmed this with our survey but identified interactions of age and gender that may bear on the future of the competing variants. Similarly, for the <c> in *grocery*, we can account for the occurrences of both /s/ and /ʃ/ in Kansas City in Vaux & Golder’s (2003) map, but identify /ʃ/ as winning out in apparent time.

We do not differentiate between /ɑ/ and /ɔ/ in the second syllable of *crayon*, but like Vaux & Golder (2003), we find a strong preference for disyllabic pronunciations in Kansas City. However, we find 26.5 percent of Kansas Citians selecting a monosyllabic /ɑʊ/ and 12.0 percent selecting monosyllabic /æ/, which suggests a much greater presence for these variants than the one /æ/ dot plotted in the Harvard map. Similarly, for the name of the freshwater lobster-like creature, Vaux & Golder’s (2003) map plots a preference for *crawdad* in Kansas City and limited occurrences of *crawfish* and *crayfish*.

We have accounted for all three variants and identified interactions with race and gender.

We are unaware of research on the other lexical items we surveyed in Kansas City. For these we find a racial split between *teeter-totter* and *seesaw*, erasure of the *Bootheel/Boot Hill* of Missouri for people who claim Kansas City as their hometown, a change in apparent time that is replacing *anyway* with *anyways*, and *Highway 24/24 Highway* splitting across the Missouri River. We find many competing variants for the term for someone who thinks they are better than everyone else, but no pattern for social factors.

Among the grammar items included in the MLS, previous research allows us to posit some hypotheses for positive-*anymore*. Kansas Citians would have almost certainly been included among the Missourians that Youmans (1986:69) surveyed, for whom he reports acceptability rates of 35- to 48-percent acceptability rates across two surveys for the item “Those are worthless *anymore*”, and 38- to 40-percent acceptability for “*Anymore* those are worthless”. Four positive-*anymore* items in Vaux & Golder’s (2003) maps suggest that a preponderance of informants reject the feature, while Ash’s (2006: 50) map shows several responses of “can say” near Kansas City for the phrase “Cars sure are expensive *anymore*!”. We report rejection rates (i.e., informants claiming never to have heard the feature) for positive-*anymore* in both clause-initial and -final position, at 44.7 and 35.7 percent, respectively. We cannot directly compare MLS responses to the acceptability judgments in Youmans (1986) or Vaux & Golder (2003). Nor can we make much of the distinction between responses of “use” and “heard”, since it is well established that speakers do not report reliably on their use of positive-*anymore* (e.g., Labov 1972: 309) (though we do see similar rates for responses of “heard” between the two instances of positive-*anymore* and the control negative-*anymore*, suggesting that we might treat both “use” and “heard” as offering insight into acceptability). Nevertheless, we confirm the presence of positive-*anymore* in Kansas City in a large dataset, while also showing evidence against increasing acceptability of positive-*anymore* in apparent time. We show differentiation in acceptability of positive-*anymore* according to race. We are unfamiliar with prior research that examines positive-*anymore* for interactions with race.

Like *positive-anymore*, the *needs*+PAST PARTICIPLE construction is difficult to study by speaker intuition. Ash (2006: 48-49) reports its presence in Kansas City. We offer strong confirmation for this finding—indeed, informants report using the form at greater rates than they do the control *negative-anymore* (55.3 percent versus 49.1 percent). We also find a race effect that connects *needs*+PAST PARTICIPLE particularly with white informants. Bloomquist's (2009: 38-39) study of central Pennsylvania is the only work we are aware of that examines race as a predictor for acceptability judgments of *needs*+PAST PARTICIPLE. However, she found no significant differences between white and African American respondents for the variable. So our findings suggest a different racial dynamic for *needs*+PAST PARTICIPLE in Kansas City than she found in Pennsylvania.

Race is also a factor in responses to *want*+PREPOSITION ADVERB, with African Americans reporting less use of the construction than other informants. We strongly confirm the presence of this Midland grammatical feature in Kansas City (cf. Murray & Simon 2006: 26-27; Benson 2012: 226) but offer race as a novel explanatory factor.

Ash (2006:49) counts *come with* \emptyset in the phrase "Do you want to *come with*?" as occurring in Kansas City. However, the map (Ash 2006: 52) shows primarily responses of "not heard", so it is unclear how she reaches this interpretation. Vaux & Golder (2003) plot conflicting acceptability responses to the construction in Kansas City. By contrast, we identify the construction as a vigorous change in progress in Kansas City, with younger (and more urban) Kansas Citians rapidly adopting the usage.

Punctual-*whenever* has been noted as a feature of Midland US English (Montgomery & Kirk 2001; Murray & Simon 2006: 27-28), but not studied in Kansas City. We confirm its presence in Kansas City, and our statistical measures suggest that it is relatively stable (i.e., not advancing).

4.2 Goal 2: Explore the Role of Borders in the Community's Language Practices

While borders crisscross Kansas City, from a linguistic perspective the border that matters most is race. In a sense, this is a trivial observation; very broadly speaking, across the United States the Englishes of African Americans and whites are diverging

(e.g., Labov 2012). Given the deeply segregated history of Kansas City and the continued impenetrability of longstanding racial divisions, it is unsurprising that Kansas City mimics the broader US pattern of language divergence between whites and African Americans.

Nevertheless, the negative results for all other sorts of social borders—especially geopolitical borders of city, county, state, and the Missouri River—is noteworthy. We are not able to comment strongly on the Missouri-Kansas border, since our sample is limited to seventeen informants explicitly from Kansas and potentially a few more who listed “Kansas City” as their hometown, but we do have sufficiently large numbers of informants to reveal differing linguistic patterns that might arise from other very meaningful geopolitical divisions. While many adults will of course interact socially and professionally across physical and political boundaries so that the solidity of these borders is doubtable, adolescents and teenagers in their formative linguistic years naturally have more restricted social networks that will often conform to lines on maps. It is not much more likely, for instance, that a teenager in a rural high school would interact substantively with a teenager in a suburban high school (or a teenager north of the Missouri River with one to the south) than it is that a white teenager would interact meaningfully with an African American teenager. As such, there is an impetus in place for linguistic differences to emerge according to geopolitical boundaries besides race.

However, such differences do not emerge in responses to the MLS. With a very few qualified exceptions—*Highway 24/24 Highway*, *Bootheel/Hill*, and *come with Ø*—city, county, state, and river do not predict responses to survey items. This lack of difference is true across the entire sample and when whites and African Americans are modeled in isolation. Regardless of other geopolitical or psychological boundaries, then, Kansas Citians talk like members of their racial groups.

This result is, of course, limited to conscious knowledge about language derived from a written questionnaire. Other methodologies may well reveal more nuanced linguistic differences. But for the present research, however real the physical, political, and psychological borders in Kansas City are, the linguistically important barrier is race.

4.3 Goal 3: Explore the Cohesiveness of the Speech Community

While we cannot speak in any definitive way about the nature of the “speech community” from the present study’s methods and results, to the extent that our findings identify race as a meaningful demarcation in Kansas City’s linguistic practices and deny the importance of other geopolitical boundaries, we can at least offer some support for the idea of Kansas City as a speech community. In consciously accessible language judgments, “place” within Kansas City does not bear much on the conclusions we draw about language in Kansas City.

We offer this as tentative support to the idea that it is possible to establish a broad baseline for language practice within a community, which can in turn be used for closer studies of networks, groups, and individuals. It potentially admits a much less rigorous sampling method to establish a community baseline than the sociological model that Labov (2001: 39) demands. To the extent that researchers in any “wave” of variationist sociolinguistics need to draw on conscious judgments about language to create a set of norms for a community, we provide an example of a large metropolitan area that we can speak of as a speech community. MLS responses suggest that—as long as our sample is racially homogenous—we could survey a single city, sample widely across the metropolitan area, or work intensively with a fairly small group, and still talk meaningfully about English in “Kansas City”.

Of course, we offer this suggestion very hesitantly. Strelluf (2018) includes details on Kansas Citians’ family and social networks that could be compared against broader linguistic patterns and practices in the metropolitan area, but does not offer any such comparison. A focused comparison of productions in micro communities of practice against macro communities of practice might offer a much more informative conclusion about the nature of Kansas City as a speech community and the “speech community” as a unit of analysis in sociolinguistics. Nevertheless, we suggest that MLS data collected on Kansas City—beyond providing a dialect topography of the community for a number of previously unstudied linguistic variables—also offers ways to understand the community itself and to contemplate the nature of communities in sociolinguistic studies.

References

- AGHA, Asif (2003) "The social life of a cultural value", *Language and Communication* 23(2), 231-273.
- ASH, Sharon (2006) "The North American Midland as a dialect area", in Thomas E. Murray & Beth Lee Simon (eds.), *Language variation and change in the American Midland*, Amsterdam: John Benjamins, 33-56.
- BENSON, Erica J. (2012) "Need + prepositional adverb in the Midland: Another feature *needs in*", *Journal of English Linguistics* 40(3), 224-255.
- BLOOMQUIST, Jennifer (2009) "Dialect differences in central Pennsylvania: Regional dialect use and adaptation by African Americans in the lower Susquehanna Valley", *American Speech* 84(1), 27-47.
- BOTTOMSUPCOMEDY (2013) A tour of Kansas City. YouTube. <<https://www.youtube.com/watch?v=x2ZCddp4bIY>> (Accessed June 19. 2017).
- BROWN, A. Theodore & Lyle W. DORSETT (1978) *K.C.: A history of Kansas City, Missouri*, vol. 2, Western Urban History Series, Boulder: Pruett.
- BUCHOLTZ, Mary (1999) "'Why be normal?': Language and identity practices in a community of nerd girls", *Language in Society* 28(2), 203-223.
- CARVER, Craig (1987) *American regional dialects*, Ann Arbor: University of Michigan Press.
- CHAMBERS, J. K. (1994) "An introduction to dialect topography", *English World-Wide* 15(1), 35-53.
- CHAMBERS, J. K. (1998) "Inferring data from a postal questionnaire", *Journal of English Linguistics* 26(3), 222-246.
- CHAMBERS, J. K. & Peter TRUDGILL (1998) *Dialectology*, 2nd edition, Cambridge: Cambridge University Press.
- CHEVALIER, Alida (2016) *Globalisation versus internal development: The reverse short front vowel shift in South African English*, Unpublished PhD Thesis: University of Cape Town.
- DOLLINGER, Stefan (2015) *The Written questionnaire in social dialectology: History, theory, practice* (Studies in Language and Society 40), Philadelphia: John Benjamins.
- ECKERT, Penelope (2004) *Vowel shifts in Northern California and the Detroit suburbs: Northern California vowels*. <<http://www.stanford.edu/~eckert/vowels.html>> (30 November 2013)
- ECKERT, Penelope (2012) "Three waves of variation study: The Emergence of meaning in the study of sociolinguistic variation", *Annual Review of Anthropology* 41, 87-100.
- ECKERT, Penelope & Sally McCONNELL-GINET (1992) "Think practically and look locally: Language

- and gender as community-based practice”, *Annual Review of Anthropology* 21(3), 461-490.
- FARIES, Rachel Bernice (1954) *A Survey of the vocabulary of seven Northeast Central Missouri Counties*, University of Missouri MA thesis.
- FARIES, Rachel Bernice (1967) *A Word geography of Missouri*, University of Missouri PhD dissertation.
- FARIES, Rachel B. & Donald M. LANCE (1993) “Regional Variation in Missouri”, in Timothy C. Frazer (ed.), *“Heartland” English: Variation and transition in the American Midwest*, Tuscaloosa: University of Alabama Press, 245-256.
- FRIDLAND, Valerie, Tyler KENDALL, Betsy E. EVANS & Alicia Beckford WASSINK (2016) *Speech in the western states*, vol. 1, *The coastal states* (Publication of the American Dialect Society 101), Durham, NC: Duke University Press.
- GORDON, Elizabeth, Lyle CAMPBELL, Jennifer HAY, Margaret MACLAGAN, Andrea SUDBURY & Peter TRUDGILL (2004) *New Zealand English: Its origins and evolution*, Cambridge: Cambridge University Press.
- GORDON, Matthew J. (2006) “Tracking the low back merger in Missouri”, in Thomas E. Murray & Beth Lee Simon (eds.), *Language variation and change in the American Midland*, Amsterdam: John Benjamins, 57-68.
- HOLMES, Janet & Miriam MEYERHOFF (1999) “The community of practice: Theories and methodologies in language and gender research”, *Language in Society* 28(2), 173-183.
- HOTHORN, Torsten, Peter BUEHLMANN, Sandrine DUDOIT, Annette MOLINARO & Mark Van Der LAAN (2006) “Survival ensembles”, *Biostatistics* 7(3). 355-373.
- HOTHORN, Torsten, Kurt HORNIK, Carolin STROBL & Achim ZEILEIS (2017) *party: A Laboratory for recursive partytioning*. Ver. 1.2-3. <<http://party.r-forge.r-project.org/>>.
- HOTHORN, Torsten, Kurt HORNIK & Achim ZEILEIS (2006) “Unbiased recursive partitioning: A Conditional inference framework”, *Journal of Computational and Graphical Statistics* 15(3), 651-674.
- JOHNSTONE, Barbara (2009) “Pittsburghese shirts: Commodification and the enregisterment of an urban dialect”, *American Speech* 84(2), 157-175.
- “Kansas City Barbie” (2008) eBaum’s World. <<http://www.ebaumsworld.com/pictures/kansas-city-barbie/292485/>> (July 11, 2017).
- KENNEDY, Robert & James GRAMA (2012) “Chain shifting and centralization in California vowels: An acoustic analysis”, *American Speech* 87(1), 39-56.

- KIESLING, Scott (1998) "Men's identities and sociolinguistic variation: the case of fraternity men", *Journal of Sociolinguistics* 2(1), 69-100.
- LABOV, William (1966/2006) *The social stratification of English in New York City*, 2nd edition, Cambridge: Cambridge University Press.
- LABOV, William (1972) *Sociolinguistic patterns*, Philadelphia: University of Pennsylvania Press.
- LABOV, William (2001) *Principles of linguistic change*, vol. 2, Social factors, Oxford: Blackwell.
- LABOV, William (2012) *Dialect diversity in America: The politics of language change*. Charlottesville, VA: University of Virginia Press.
- LABOV, William, Sharon ASH & Charles BOBERG (2006) *Atlas of North American English: Phonetics, phonology, and sound change*, Berlin: Mouton de Gruyter.
- LANCE, Donald M. & Rachel B. FARIES (1997) "Regional vocabulary in Missouri", in Cynthia Bernstein, Thomas Nunnally & Robin Sabino (eds.), *Language variety in the South revisited*, Tuscaloosa: University of Alabama Press, 361-381.
- LUSK, Melanie M. (1976) *Phonological variation in Kansas City: A Sociolinguistic analysis of three-generation families*, PhD dissertation, University of Kansas.
- MARC (MID-AMERICA REGIONAL COUNCIL) (2017) *Population estimates by age, sex, race, and Hispanic origin*.
<<http://www.marc.org/Data-Economy/Metrodateline/Population/Current-Population-Data>> (July 6, 2017).
- MEYERHOFF, Miriam (2006) "Communities of practice", in J.K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, Oxford: Blackwell, 526-548.
- MILROY, Lesley & Matthew J. GORDON (2003) *Sociolinguistics: Method and interpretation* (Language in Society 34), Malden, MA: Blackwell.
- MONTGOMERY, Michael B. & John M. KIRK (2001) "'My mother, whenever she passed away, she had pneumonia': The history and functions of whenever", *Journal of English Linguistics* 29(3), 234-249.
- MOORE, Emma & Paul CARTER (forthcoming) "Natural phonetic tendencies and social meaning: Exploring the allophonic raising split of PRICE and MOUTH on the Isles of Scilly".
- MURRAY, Thomas E. & Beth Lee SIMON (eds.) (2006) *Language variation and change in the American Midland*, Philadelphia: John Benjamins.
- O'HIGGINS, Briana (2014) "How Troost became a major divide in Kansas City", *Beyond Our Borders*. <<http://kcur.org/post/how-troost-became-major-divide-kansas-city>> (July 2, 2014).

- PODESVA, Robert J. (2011) "The California Vowel Shift and gay identity", *American Speech* 86(1), 32-51. doi:10.1215/00031283-1277501.
- R CORE TEAM (2016) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
- SHORTRIDGE, James R. (2012) *Kansas City and how it grew, 1822-2011*, Lawrence, KS: University Press of Kansas.
- STEPHENS, Trevor (2014) Titanic: Getting started with R. Trevor Stephens. <<http://trevorstephens.com/kaggle-titanic-tutorial/getting-started-with-r/>> (June 4, 2017).
- STRASSER, Helmut & Christian WEBER (1999) "On the asymptotic theory of permutation statistics", *Mathematical Methods of Statistics* 8(2), 220-250.
- STRELLUF, Christopher (2016) "Overlap among back vowels before /l/ in Kansas City", *Language Variation and Change* 28(3), 379-407. doi:10.1017/S0954394516000144.
- STRELLUF, Christopher (2018) *Speaking from the heartland: The Midland vowel system of Kansas City* (Publication of the American Dialect Society 103), Durham, NC: Duke University Press.
- STROBL, Carolin, Anne-Laure BOULESTEIX, Thomas KNEIB, Thomas AUGUSTIN & Achim ZEILEIS (2008) "Conditional variable importance for random forests", *BMC Bioinformatics* 9(307) <<http://www.biomedcentral.com/1471-2105/9/307>>
- STROBL, Carolin, Anne-Laure BOULESTEIX, Achim ZEILEIS & Torsten HOTHORN (2007) "Bias in random forest variable importance measures: Illustrations, sources and a solution", *BMC Bioinformatics* 8(25) <<http://www.biomedcentral.com/1471-2105/8/25>>.
- TAGLIAMONTE, Sali A. & Harald R. BAAYEN (2012) "Models, forests, and trees of York English: was/were variation as a case study for statistical practice", *Language Variation and Change* 24(2), 135-178.
- "The New border war" (2014) *The Economist*. Mar. 22 <<http://www.economist.com/news/united-states/21599368-missouri-calls-economic-truce-kansas-new-border-war>> (June 20, 2017).
- US CENSUS (2013) *Annual estimates of the resident population: April 1, 2010 to July 1, 2013—United States—Metropolitan Statistical Area; and for Puerto Rico*. <<http://factfinder2.census.gov/>> (July 2, 2014).
- US CENSUS (2015) *Metropolitan and Micropolitan Statistical Areas (CBSAs) of the United States and Puerto Rico: July 2015*.

http://www2.census.gov/geo/maps/metroarea/us_wall/Jul2015/cbsa_us_0715.pdf

(November 11, 2016).

US CENSUS (2016) *Quick facts*, July 1, 2016. <<https://www.census.gov/quickfacts/>> (July 6, 2017).

VAUX, Bert & Scott GOLDER (2003) *The Harvard dialect survey*, Cambridge, MA: Harvard University Linguistics Department.

WORLEY, William S. (1990) *J.C. Nichols and the shaping of Kansas City: Innovation in planned residential communities*, Columbia, MO: University of Missouri Press.

YOUNG, Gilbert (1986) "Anymore on *anymore*?: Evidence from a Missouri dialect survey", *American Speech* 61(1), 61-75.

ZELINSKY, Wilbur (1992) *The cultural geography of the United States*, revised edition, Englewood Cliffs, NJ: Prentice Hall.

Appendix A: Missouri Language Survey

This survey is part of a research project studying speech patterns in Missouri. If you choose to participate, it should take approximately 10 minutes for you to fill out the survey. You are asked to provide some background information about yourself but not your name, so your identity won't be known to researchers. Your participation is voluntary, and you can choose to withdraw from the study by not filling in the survey or not returning it to the student researcher. If you have ever completed this survey before, please do not complete it again. If you have any questions about this study, you can ask the student researcher or contact Prof. Christopher Strelluf of Northwest Missouri State University (cstrell@nwmissouri.edu).

Please answer these questions about yourself:

Age: _____ Sex: M / F Ethnic Background: _____
Hometown (or place you lived during high school): _____

Please list any other places you have lived and indicate when in your life you lived in each place (e.g., Independence birth –age 4, Hannibal age 5 – 9, Branson age 10 – 17, etc.):

Please answer these questions about your way of speaking:

1. Say each of the following pairs of words aloud using your normal, everyday pronunciation. Then, indicate whether the words sound **(1)** exactly the same; **(2)** similar but distinct; **(3)** very different from each other; or **(4)** you can say them either the same or distinct. Please circle your choice.

| <u>Same</u> | | | <u>Close</u> | <u>Different</u> | <u>Either</u> |
|-------------|----------------------------|----------|--------------|------------------|---------------|
| a. | <i>wait</i> <i>weight</i> | 1 | 2 | 3 | 4 |
| b. | <i>haul</i> <i>hull</i> | 1 | 2 | 3 | 4 |
| c. | <i>caught</i> <i>coat</i> | 1 | 2 | 3 | 4 |
| d. | <i>cot</i> <i>caught</i> | 1 | 2 | 3 | 4 |
| e. | <i>gym</i> <i>gem</i> | 1 | 2 | 3 | 4 |
| f. | <i>whole</i> <i>hole</i> | 1 | 2 | 3 | 4 |
| g. | <i>pull</i> <i>pool</i> | 1 | 2 | 3 | 4 |
| h. | <i>dawn</i> <i>Don</i> | 1 | 2 | 3 | 4 |
| i. | <i>pen</i> <i>pin</i> | 1 | 2 | 3 | 4 |
| j. | <i>Polly</i> <i>Paulie</i> | 1 | 2 | 3 | 4 |
| k. | <i>goal</i> <i>gull</i> | 1 | 2 | 3 | 4 |
| l. | <i>here</i> <i>hear</i> | 1 | 2 | 3 | 4 |
| m. | <i>tour</i> <i>tore</i> | 1 | 2 | 3 | 4 |
| n. | <i>bull</i> <i>bowl</i> | 1 | 2 | 3 | 4 |

2. Which of the following words have the same vowel sound as **bought**? (please circle all that apply)

caught lot dog hawk box what odd cough rough hot

3. Which of the following words have the same vowel sound as **tin**? (please circle all that apply)

sin seen pen hid wet tent sit

4. Which of the following words have the same vowel sound as **full**? (please circle all that apply)

foot bull cool coal call food gulf pole golf gull book pull tool

5. Does the first part of **coupon** sound like the first part of **cupid** or **coop**? (please circle your choice)

CUPID

COOP

6. Does **crayon** rhyme better with *brown* or *bran* or *play on*? (please circle)

BROWN

BRAN

PLAY ON

7. How do you normally pronounce the “c” in **grocery**? Does it sound like the “s” in *sure* or in *sir*?

SURE

SIR

8. What do you call the piece of playground equipment consisting of a long board with seats on the ends? One child sits on either end to balance it out and they go up and down.

9. How would you describe someone who thinks he’s better than everyone else? You might say:

“That person is very _____”.

10. What is the nickname for the extreme southeast corner of the state of Missouri, the part along the Mississippi River that seems to dip below the long state border into Arkansas.

11. When you refer to the numbered highways in your hometown, which word comes first: the number or “highway”? (Please circle)

For example:

24 Highway

or

Highway 24

12. Which word would you use to end this sentence: “It’s raining, but I’m going out _____”?

ANYWAYS

ANYWAY

13. What do you call the small lobster-like animal that can be found in freshwater creeks and

lakes?

14. Think about the phrasing of the sentences below and mark whether they sound like (a) phrasing that you might use in everyday speech; (b) phrasing that you have heard from other people but that you wouldn't use yourself; or (c) phrasing that you've never heard before. Try to base your choices on how the sentences are phrased and not on their content. (Please circle a, b, or c)

- | | | | |
|---|---|---|---|
| ● <i>We're going out. Do you want to come with?</i> | a | b | c |
| ● <i>Whenever we were growing up, we lived on a farm.</i> | a | b | c |
| ● <i>Food is so expensive anymore.</i> | a | b | c |
| ● <i>Kids don't get enough exercise anymore.</i> | a | b | c |
| ● <i>The boy was 10 whenever his mother died.</i> | a | b | c |
| ● <i>Anymore most things are made in China.</i> | a | b | c |
| ● <i>The car needs washed.</i> | a | b | c |
| ● <i>I want off this bus.</i> | a | b | c |