International Linguistics Research; Vol. 5, No. 2; 2022
ISSN 2576-2974 E-ISSN 2576-2982
https://doi.org/10.30560/ilr.v5n2p1

Utah Dialect Regions and Features

David Ellingson Eddington[1]

[1] Brigham Young University, Provo, Utah, USA

Correspondence: David Eddington, 4065 JFSB, Provo, Utah 84602 USA. Tel: 1-801-422-7452. E-mail: eddington @byu.edu

Received: September 7, 2022 Accepted: September 27, 2022 Online Published: September 30, 2022

Abstract

Data from the Utah Dialect Survey were used to delineate dialect boundaries in the state of Utah. Statistical algorithms divided the state into rural and urban dialect zones. Twelve features were identified that distinguish the two dialects. Three additional features were extracted from the the Harvard Survey of North American Dialects. The idea that the rural vs. urban divide is related to a greater number of immigrants to urban areas is discounted. Many rural features coincide with those used more by older Utahns suggesting that rural areas maintain more conservative Utah characteristics.

Keywords: dialect geography, Utah dialect, dialectometry, American English

1. Introduction

The study of dialectal features in the state of Utah most likely began with Pardoe's study in 1935. However, with the advent of sociolinguistics in the later part of the 20th century the characteristics of what may be termed Utah English have attracted the attention of a number of scholars who have examined features such as the use of propredicate do (Di Paolo, 1993), vowel mergers (Bowie, 2003), and oral releasing after glottal stops (Eddington & Savage, 2012). The popular press, for its part, has done much to propagate stereotypes along with misinformation about the speech of Utahns (e.g. DeBry, 2019), and residents of the state have some unfavorable attitudes about certain features that are prominent in Utah (Savage, 2014). A fruitful field of research has been how religious affiliation influences speech patterns in Utah (Baker-Smemoe & Bowie, 2015), as well as how some features may be due to the large numbers of Danish (Graham, 2006) and English immigrants (DiPaolo, 1993) to the state.

Conversation about dialects among Utah residents often focuses on the perception that people in one part of the state stand out, or that there is a particular dialect in a given region. The purpose of the present paper is investigate this idea more systematically by attempting to determine geographic dialect boundaries. This is a challenging task since we know that isoglosses for different features are notorious for not coinciding with each other, except where topographic features such as oceans and mountain ranges are found. What is more, the use of isoglosses is also suspect since linguistic features are rarely binary, as isoglosses suggest, but scalar in nature. Often, the differences between the dialect boundaries proposed by different researchers is principally the result of the varying criteria employed by each one.

The data available on the internet, the widespread availability of powerful computers, and the existence of large corpora have led to innovative approaches to dialect studies (Grieve et al., 2011). The most notable characteristic of contemporary approaches is that they do not depend on small numbers of features, but follow the advice of researchers who argue that dialectology must aggregate large numbers of features to obtain maximally reliable results (e.g. Nerbonne, 2009; Séguy, 1971). Allowing a computer algorithm, rather than the researcher, to determine the boundaries also eliminates one source of subjectivity. The first goal of the present study is to determine dialect regions in Utah. This is accomplished by statistical means. The data the analyses are based on come from a recent dialect survey of the state (Eddington, 2023). The second goal of the paper is to examine the features that distinguish the resulting dialect areas.

2. Previous Dialect Surveys of Utah

A number of studies of regional variation in the US have gathered data from Utah. The Atlas of North American English (Labov, Ash, & Boberg, 2008) only included data from seven Utahns, while Vaux and Golder's survey (2003) garnered responses from 322. However, neither of these were designed to examine in-state dialectal differences, although later in the paper Vaux and Golder's Utah data are used to this end.

In contrast, Lillie (1998) discusses the results of her study in terms of the tripartite division of the state illustrated in Figure 1. She carried out her dialect survey with the help of students from three universities who administered the survey in person to 732 Utahns. Her survey examined 42 features such as vowel mergers before sonorants (e.g. *cord/card, feel/fill, pin/pen, fail/fell*), euphemistic swearing, and vocabulary variation (e.g. *pop/soda, you bet* for *thank you, stop light/traffic light*). Her major findings for each area are summarized on her web page (Lillie n.d.).

Figure 1. Lillie's division of Utah into three dialect zones

3. Utah Dialect Survey

In order to build on, and update Lillie's survey, the Utah English Survey was carried out (Eddington, 2023) online in 2020, and obtained responses from 1764 Utahns from all counties. In the survey, 28 dialect features were tested (Table 1) in a series of 88 questions. The influence of social variables such as age, religion, and gender in relation to these 28 features is discussed. Many of the 28 features were similar to those tested by Lillie, while others were chosen because they have been the topic of linguistic research on Utah English, or because they have been discussed as Utah traits in the media as well as anecdotally.

Table 1. The 28 dialect characteristics tested in the survey.

Characteristic/Phenomenon	Example
Pronounced [l]	*Balm:* [bɑm] or [bɑlm]
Fill/Feel Merger	*Meal:* m[i]l or m[ɪ]l
Fail/Fell Merger	*Jail:* j[eɪ]l or j[ɛ]l
Words Like *Measure*	*Pleasure:* pl[eɪ]sure or pl[ɛ]sure
-day in Days of the Week	*Friday:* Frid[eɪ] or Frid[i]
Creek	*Creek:* cr[i]k or cr[ɪ]k
Roof	*Roof:* r[u]f or r[ʊ]f
Cord/Card Merger	*Cord:* c[ɔ]rd or c[ɑ]rd
Pin/Pen Merger	*Pin:* p[ɪ]n or p[ɛ]n
Beg Raising	*Beg:* b[ɛ]g or b[eɪ]g
Bag Raising	*Bag:* b[æ]g, b[ɛ]g, or b[eɪ]g
Pool/Pole/Pull Merger, [ul]	*School:* sch[u]l or other vowel
Pool/Pole/Pull Merger, [ʊl]	*Full:* f[ʊ]l or other vowel
Pool/Pole/Pull Merger, [ol]	*Goal:* g[o]l or other vowel
*Hull/Hole/Hall M*erger	*Hull:* h[ʌ]l or other vowel
Words Like *Lauren*	*Lauren:* L[ɑ]ren or L[ɔ]ren
Tour	*Tour:* t[u]r or t[ɔ]r
Alternatives to *You're Welcome,* mm-hmm, or uh huh	Use *mm-hmm,* or *uh huh* sometimes or never
Alternatives to *You're Welcome, You Bet*	Use *you bet* sometimes or never
Expressions for Takeout	*To stay or to go?* Or *For here or to go?*
Expressions for Next in Line	*I'll help who's next* or *I'll help whoever's next*
Pop, Soda, or Coke	*Pop, Coke, soda, soda pop,* or *soft drink*
Route as a Highway Name	*Route:* r[u]t or r[aʊ]t
Route Meaning Path	*Route:* r[u]t or r[aʊ]t
Route as a Verb	*Route:* r[u]t or r[aʊ]t
Propredicate *Do*	How likely are you to say: *If it requires our entire life savings, A as it may do B as it may, C as it may require*
(Oh) For + Adjective	*You see something that you think is awesome, how likely are you to say the following?: How cool!, That's cool!, or For cool!* Rate how likely you are to say it.
Intrusive [t]	*Answer:* an[t]swer or answer

In the survey, questions about pronunciation asked speakers to match their pronunciation of a word with another word as in the example below:

How do you pronounce the highlighted vowel in the word bull?

(You can choose two words if you feel your pronunciation is close to both words.)

Like the word:

A *l**oo**k*

B *L**u**ke*

C *l**o**ck*

D *l**u**ck*

E *bl**o**ke*

Other questions were presented with Likert-type responses:

If you see an adorable little puppy, how likely are you say the following?

For cute!

A Very likely

B Somewhat likely

C Somewhat unlikely

D Very unlikely

All response values were converted into separate variables, and the proportion of responses for each county was calculated. For example, there were six variables for the pronunciation of *bull* which consisted of the proportion of responses in each county that paired it with *Luke, look, lock, luck, bloke* or some combination of vowels. For questions with Likert-type responses, the responses were made numeric: very likely 2, likely 1, somewhat unlikely -1, very unlikely, -2, then further reduced to a binary outcome: likely or unlikely. The mean response by county was then derived for each variable value. The resulting data set contained mean responses by county for 118 variables.

4. Determining Dialect Boundaries in Utah

The Utah Dialect Survey assumed the dialect map in Figure 2. The most densely populated counties in the state are collectively known as the Wasatch Front, which runs along the western slope of the Wasatch mountain range. However, it should be noted that there is no consensus on exactly which counties the Wasatch Front comprises (Wikipedia, 2021; Wikitravel, 2022). However, the counties highlighted in dark gray were called the limited Wasatch Front, and the when four contiguous counties in light gray are added to the limited Wasatch Front that region was referred to as the extended Wasatch Front. The remainder of the state was considered a single dialect area.

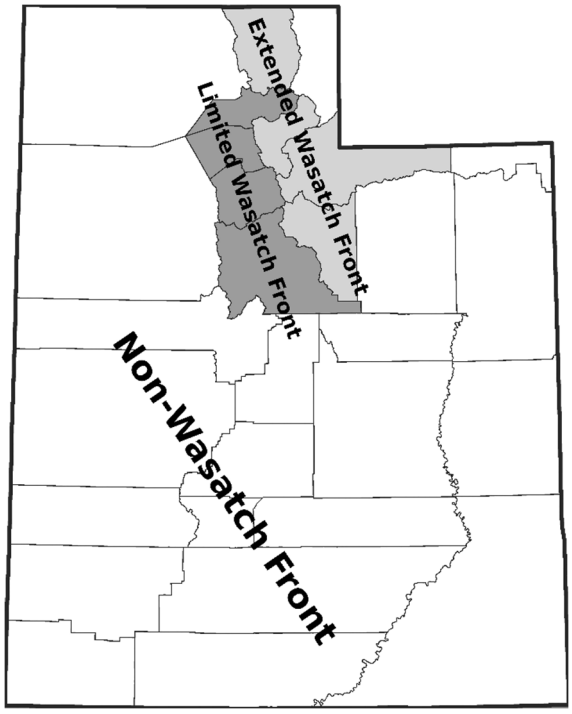

Figure 2. Utah Dialect Survey's division of Utah into dialect zones

In the statistical analysis of the survey, a number of variables were included that related to geographical dialect areas: individual county, county population, the tripartite division proposed by Lillie (1998), the limited Wasatch Front versus other counties combined, the extended Wasatch Front versus other counties combined. In the analysis of the survey responses, several of the geographical variables were significant. The best fitting geographical variable was determined by comparing the resulting R2 values for each geographical division and choosing the division that accounted for more of the variance. Dividing the state into thirds, as Lillie suggested, resulted significant in the analysis of six traits. The limited Wasatch Front was significant in seven, and the extended Wasatch Front in ten. The difficulty with the divisions suggested so far is that the counties they comprise were all delineated a priori. Therefore, the question of what counties cluster together needs to be determined exclusively on dialect data, rather than on preconceived notions, which is the purpose of the present study.

As described above, the mean responses from the 2020 dialect survey for each county were calculated. However, the counties were divided in two ways. The first was based on the county the participants were raised in, and the second on the county they resided in. Responses for participants who were not raised in Utah or were raised in more than one county were eliminated in the county raised data set. In like manner, participants who currently did not reside in Utah were deleted from the county of residence data set. There were no residents in Piute county so

those data were imputed from the responses that corresponded to the county the participants were raised in.

The raised and reside data sets were examined separately. First, counties were grouped using a hierarchical clustering dendogram (Seol 2020) using the jamovi software package (The jamovi Project, 2021). The Canberra distance measure was chosen because it handles data containing large numbers of zeros (Desagulier, 2014: 163), and is sensitive to small changes (Everitt et al., 2011:50). The Ward D2 clustering method was applied to the task since it minimizes the amount of variance, produces compact clusters, and is commonly used in linguistic tasks (e.g. Grieve et al., 2011; Gries, 2010; Prokic & Nerbonne 2008). The dendogram results in a structure similar to a family tree except that it may contain more than two branches. The branches farther down the tree incorporate fewer clusters of related data, while a the same time they represent more granularity than higher branches. The algorithm was asked to make clusters of counties into two, three, and four groups.

Next, the county data were clustered using a fuzzy C means algorithm (Bezdek et al., 1984) which works by clustering counties that are as similar as possible together while at the same time contrasting them with other clusters of counties that are as dissimilar as possible. This was done without specifying how many clusters to divide the data into, and was carried out in JASP (JASP Team, 2022). It resulted in a two-way split for both the county the participant was raised in and the participant's residence.

The dendogram and fuzzy C analyses produced four groupings by county of residence and four by county raised in. Rather than choosing a priori whether to use the data based on what county the participants were raised in or which one they resided in, and rather than subjectively choosing one of the eight groupings, the number of times each county was paired with every other county was tabulated in the eight groupings. A fuzzy C clustering algorithm was then used on the resulting data table of county pairings and it yielded the binary split illustrated in the map in Figure 3.

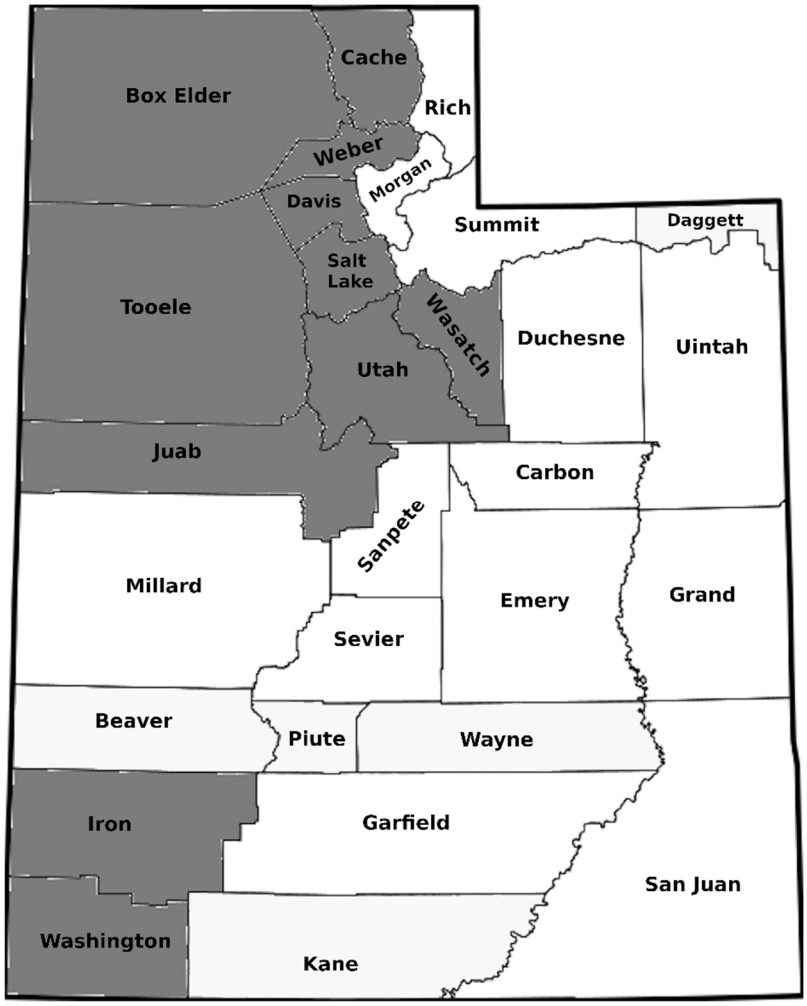

Figure 3. Computationally-derived dialect map based on survey data

The resulting groupings of counties do not coincide with Lillie's (1998) division of the state into northern, central, and southern regions, nor does it correspond to what was previously referred to either as the limited or extended Wasatch Front. Instead, the two dialect areas are primarily related to the population of the counties. The gray counties will be referred to as urban, and the white counties as rural. The number of participants who belong to the resulting rural and urban dialects appears in Table 2

Table 2. Number of Participants and Where They Were Raised or Reside.

	Urban	Rural	Many Counties	Outside of State
Raised	898	462	76	328
Reside	588	446	--	58

Juab and Wasatch Counties are somewhat exceptional since they have small populations and are surrounded by rural counties according to population (Figure 4). Nevertheless, they cluster with the urban counties linguistically. Including Juab County in the urban group makes sense since half of the residents of Juab county reside in Nephi, which is only a 19 mile freeway drive from Santaquin, the southernmost town in Utah County. In like manner, although the county seat of Wasatch County, Heber City, only has a population of 15,000, it lies a mere 28 miles from Provo in populated Utah County. The two urban southern counties that are geographically separated from the northern urban counties are Iron and Washington Counties. In Iron County, 58% of the 55,000 Utahns who call it home live in Cedar City, a college town located on Interstate 15, while Washington County houses the largest population (178,000) outside of the Wasatch Front.

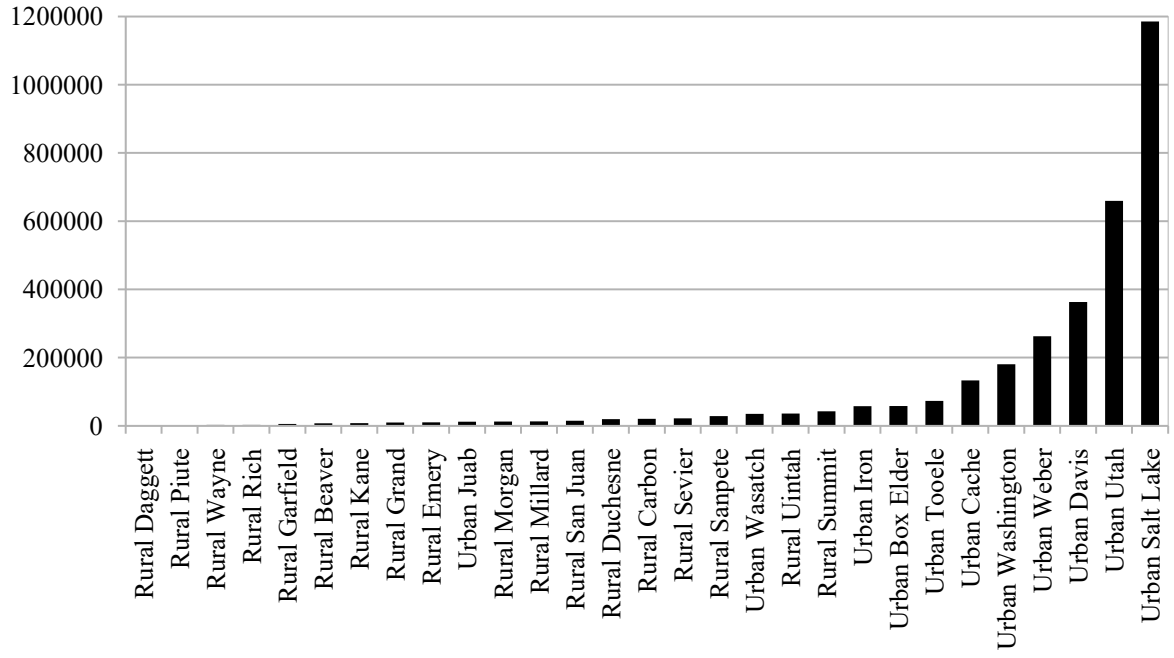

Figure 4. Counties by rural/urban division and county population

In the Utah English Survey, participants were only classified according to their county, which means that more fine-grained geographical locations were not possible. However, the shaded areas of the map in Figure 5 shows those regions with a population of 100 people per square mile or greater according to the 2010 US Census. As mentioned above, the most populated areas of Box Elder, Tooele, Juab, Wasatch, and Summit counties are located closest to the Wasatch Front which suggests a possible further subdivision of those counties in which the populated areas belong to the urban Utah dialect and the less populated areas into the rural dialect. Only further research will prove if this division is warranted or not.

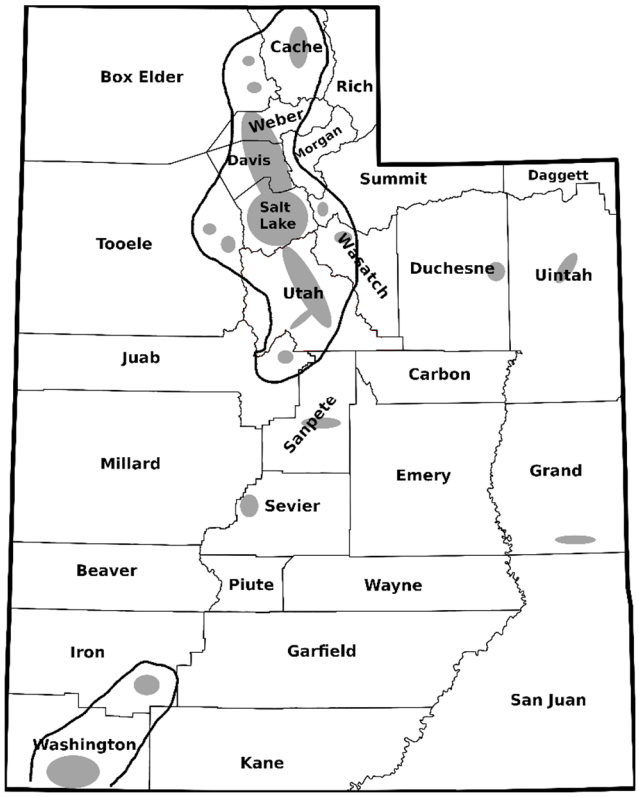

Figure 5. Population density and proposed dialect boundaries

5. Characteristics of Urban and Rural Utah Dialects

The clustering algorithms do not make clear what features they use in making their groupings. What is more, they used the county proportion of each response for each county, rather than individual responses by each participant. Therefore it is of interest to determine what dialect features are related to the rural and urban groups.

In the Utah Dialect Survey, it was necessary to limit the number of test items to make it possible to complete the survey in about ten minutes. What this means is that not all survey items were responded to by all participants. For example, to test the merger between [i] and [ɪ] some participants were asked about their pronunciation of *meal*, while others responded to *deal*. This made it necessary to analyze the data from the two parts of the survey in different binomial logistic regressions. Each of the two survey halves were then analyzed according to the region (urban or rural) that the participants were raised in and resided in, which entailed four separate analyses. Given the large number of variables, the most parsimonious model containing only variables that reached significance is reported. The reference level in all the analyses was rural. The statistical results of the data (Tables 3–6) reveal 12 features in which the rural and urban dialect regions differ. Although the results are discussed in terms of percentage differences, the odds ratio may be consulted as a measure of effect size.

Table 3. Logistic regression results: Raised in rural versus urban region, survey half one.

Predictor	Estimate	Z	p	Odds ratio	% Urban	A % A Rural	%B Urban	%B Rural
Measure: A [ɛ], B [eɪ]	0.65	2.68	.007	1.90	88	77	12	23
Pop: A *soda* B *pop*	0.78	3.66	< .001	2.18	69	48	19	38
For cool: A Unlikely to say B Likely to say	0.84	2.66	.008	2.32	95	86	5	14
Laura: A Use named person's own pronunciation, B [ɑ]	0.91	3.98	< .001	2.48	35	20	52	69
You bet for *You're welcome*: A Never, B sometimes	0.43	2.12	.033	1.54	40	29	60	71
Pool: A [u], B [ʊ]	0.63	2.97	.003	1.88	60	45	24	36

McFadden's $R^2 = .111$

Table 4. Logistic regression results: Raised in rural versus urban region, survey half two.

Predictor	Estimate	Z	p	Odds ratio	% A Urban	% Rural	A %B Urban	%B Rural
Pop: A *soda* B *pop*	1.23	5.38	< .001	3.41	71	48	17	38
Full: A combo B [ʊ]	1.20	3.25	.001	3.31	20	7	26	29
Full: A combo, B [ʌ]	0.97	2.67	.008	2.62	20	7	33	32
Full: A combo, B [o]	1.18	2.74	.006	3.24	20	7	12	13
Full: A combo, B [u]	1.38	3.27	.001	3.96	20	7	9	18
Stalk: A [stɑk], B [stɑlk]	0.57	2.09	.036	1.77	88	80	12	20
Measure: A [ɛ], B [eɪ]	0.73	2.97	.003	2.07	86	72	14	28
Cool: A [u], B [ʊ]	0.65	2.88	.004	1.92	61	50	24	37

McFadden's $R^2 = .123$

Table 5. Logistic regression results: Reside in rural versus urban region, survey half one.

Predictor	Estimate	Z	p	Odds ratio	% A Urban	% A Rural	%B Urban	%B Rural
Across: A no t, B intrusive t	2.44	2.67	.008	11.46	88	64	12	32
Pop: A *soda* B *pop*	2.68	2.77	.006	14.51	71	53	18	36
Measure: A [ɛ], B [eɪ]	2.45	2.69	.007	11.58	87	76	13	24

McFadden's $R^2 = .316$

Table 6. Logistic regression results: Reside in rural versus urban region, survey half two.

Predictor	Estimate	Z	p	Odds ratio	% A Urban	% A Rural	%B Urban	%B Rural
Pop: A *soda* B *pop*	0.75	3.50	< .001	2.11	71	54	17	32
Full: A [ʌ], B [u]	0.73	2.6	.009	2.08	34	31	10	19
Full: A [ʊ], B [u]	0.64	2.23	.026	1.89	29	28	10	19
Full: A combo, B [u]	1.36	3.77	< .001	3.89	17	8	10	19
Full: A combo, B [o]	0.9	2.33	.020	2.46	17	8	10	12
Full: A combo, B [ʊ]	0.72	2.21	.027	2.06	17	8	29	28
Lauren: A [o], B [ɑ]	0.64	2.5	.013	1.89	26	17	32	42
Meal: A [i], B [ɪ]	0.52	2.26	.024	1.69	86	77	14	23

McFadden's $R^2 = .061$

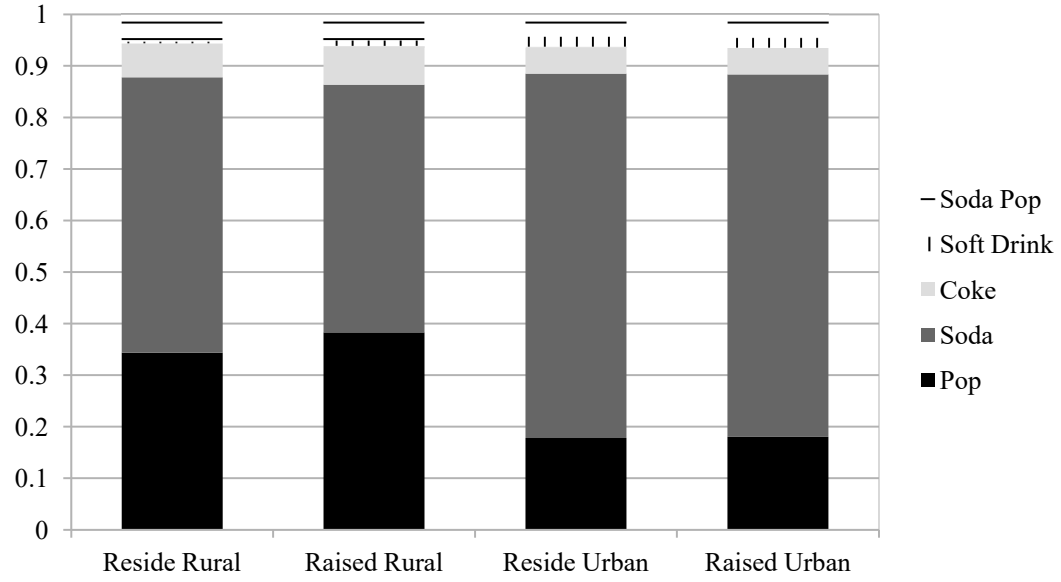

Figure 6. Terms for carbonated beverage

5.1 Carbonated Beverage

The geographic distribution of terms for carbonated beverage in the US is well documented (Abadi, 2018), and Utah appears to fall squarely into the *pop* region. However, the results of the Utah English Survey contradict those previous findings. *Soda* is the dominant term for carbonated beverages in the state. What is more, the preference for *pop* or *soda* is the single most distinguishing features of rural and urban areas (Figure 6). *Pop* is more common in rural areas. In Tables 3 to 6 it can be seen that *soda* is preferred by 69% to 88% of urban Utahns, while only 48% to 54% of rural Utahns prefer *soda.* There is, in fact, a positive correlation between county population and the percent of *soda* responses (τ (27) = .53, $p < .001$). In her 1997 survey, Lillie (1998) found that *pop* was the most frequent term in Utah. Her data also showed that *soda* was more prevalent among younger Utahns. The Utah English Survey confirms that the age-apparent shift she observed away from *pop* has taken place in real time. It appears that *pop* has been ousted by *soda* starting in the urban areas. This is most likely due to spreading of *soda* from the West Coast.

5.2 Measure: [ɛ] vs. [eɪ]

Three words ending in *-easure* appeared in the survey: *measure, treasure, pleasure.* Significant by region differences were obtained for *measure* in three of the analysis (Table 3-5). Although the most frequent pronunciation involves the vowel [ɛ], the minority [eɪ] pronunciation was preferred by about 11% more rural Utahns (Figure 7). In the survey the [eɪ] pronunciation was more prevalent among older speakers suggesting an age-apparent shift toward [ɛ] that is more advanced in urban regions. The [eɪ] pronunciation has a negative stigma attached to it in the state which was observed in a matched guise study (Savage, 2014) in which the [eɪ] pronunciation was associated with less friendly speakers.

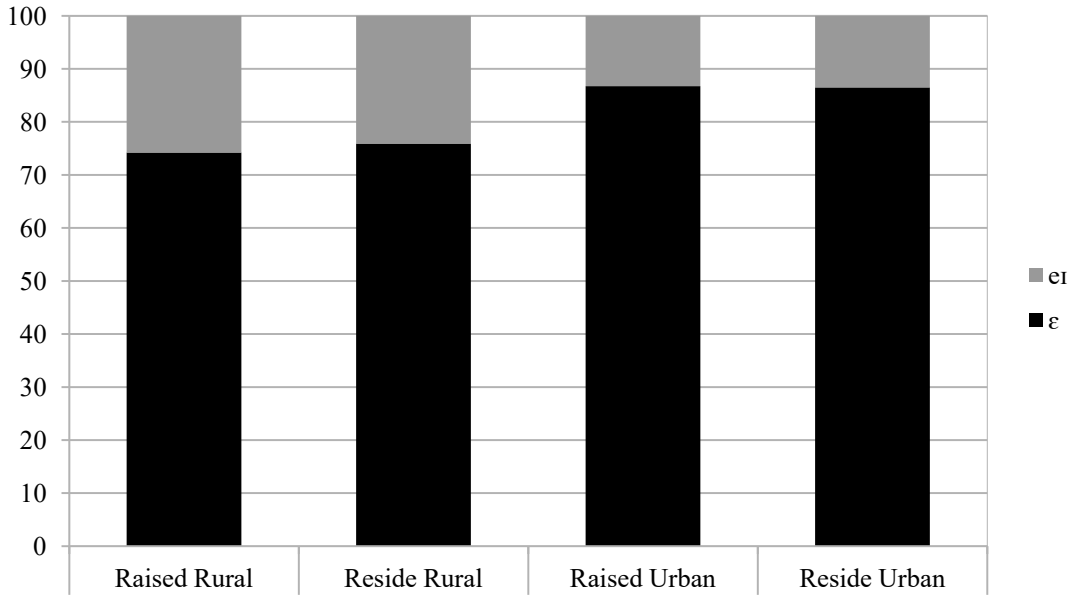

Figure 7. Preferences for [ɛ] vs. [eɪ] in *measure.*

5.3 For Cool!

Participants were asked how likely they were to use the interjection f*or cool*! Most responded that they would be unlikely, but this differed by dialect region. In Table 4, 5% of urban speakers were likely to use it compared to 14% of rural residents. In the analysis in Table 4 the percentages were a bit higher, but in the same relationship held (14% urban, 28% rural).

In American English exclamations may take the form *how* + adjective. The alternative *for* + adjective (e.g. *for cool!, for cute!*) is not only found in the Utah, but in Minnesota and Iowa as well. It it most likely due to Scandinavian influence (Graham, 2006). It appears to be a dying trait in the state since younger Utahns dispreferred it in the Utah English Survey. One thing that stands out is that acceptance of the exclamation *for cool!* was higher for residents of counties with smaller populations (τ (27) = -.482, $p < .001$).

5.4 Stalk as [stałk]

Some words vary regarding whether the orthographic 'l' is pronounced or not. The survey asked participants whether they pronounced a number of words the same or different, for example *stalk* and *stock*. The difference was assumed to be whether [ł] was pronounced in the word or not. Although a difference in vowel quality is observed in some regions of the US between these words ([ɑ] versus [ɔ]), Utah belongs to the western region of the US where the merger of these vowels renders *cot* and *caught* homophones. Significant differences by Utah dialect region were found for *stalk*, but not for *palm, caulk*, or *calm*. Only 12% of participants raised in urban areas judged *stalk* and stock to be different, while this increased to 20% in rural areas (Table 4).

5.5 You Bet for You're Welcome

While the standard response to *thank you* is *you're welcome*. Two other responses exist: *mhmm* and *you bet.* In the survey participants were asked if they used these alternatives some of the time or never. Sixty percent of Utahns raised in urban areas recognized that they sometimes say *you bet* in contrast to 71% of rural Utahns (Table 3).

5.6 Laura and Lauren as L[a]ra and L[a]ren

Perhaps the most stereotyped pronunciation in the state is the *cord/card* merger that makes homophones of *port/part* and *lord/lard* (Bowie, 2003). The merger is now generally relegated to speakers above 70 years of age (Eddington, 2023). However, the *cord/card* merger appears to have been fossilized to a certain degree in proper names such *Laura* and *Lauren*, and may be considered vestiges of this once common vowel merger. In the survey, participants were asked how they pronounced *Laura* and *Lauren* (Figure 8). 69% of the participants raised in rural areas indicated that they pronounced the name *L*[ɑ]*ra* in contrast to 52% of those raised in urban areas (Table 3). There is, in fact, a negative correlation between the pronunciation of *L*[ɑ]*ra* and the population of the county ($\tau(27) = -.582$, $p < .001$).

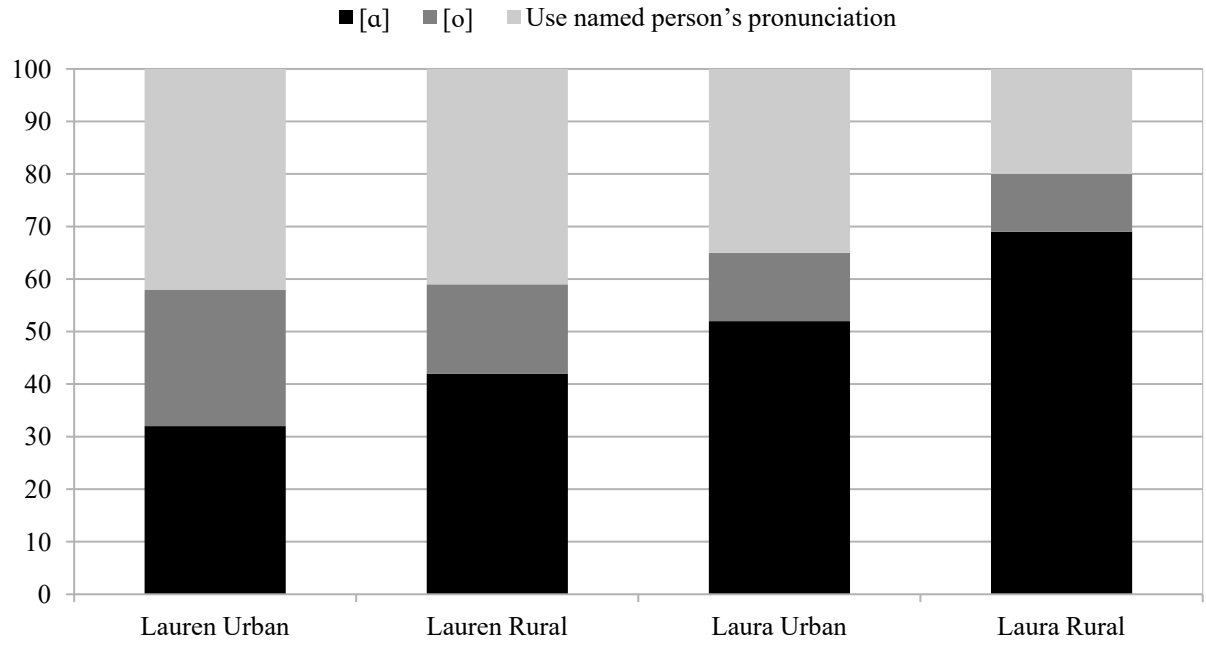

Figure 8. Pronunciation of initial vowel in *Laura* and *Lauren*

One response choice the participants were given was that they used the pronunciation of the person who carried the name. Urban Utahns were more likely (35%) to followed the pronunciation of the name used by an individual in named *Laura* in comparison to rural Utahns (20%; Table 3). As far as the name *Lauren* is concerned, 32% of urban residents preferred *L*[ɑ]*ren* over *L*[o]*ren*, while 42% of rural residents chose the [ɑ] pronunciation (Table 6).

5.7 Vowel Mergers before [ł]: full, pool, cool, meal

Some of the most commonly perceived features of Utah English have to do with the quality of vowels when they

are followed by laterals (Baker, Eddington, & Nay, 2009). In the survey, participants were asked to match the vowels in *full, cool,* and *pool* with the vowel in *look, Luke, lock, luck,* or *bloke*. If they felt that their pronunciation fell somewhere between the vowel in two or more words they were allowed to select more than one response, and these were registered as combinations. In like manner, participants matched their pronunciation of the vowel in meal with the vowel in either meet or mitt.

The responses to the word full appear in Figure 9. The principal difference is that those raised in urban Utah felt that none of the response words matched their pronunciation and hence they were more likely to choose a combination in contrast to those raise in rural areas (Table 4). Rural dwellers also disfavored combinations (Table 6), and favored the [u] vowel in full over the [ʌ] and [ʊ] pronunciations as well.

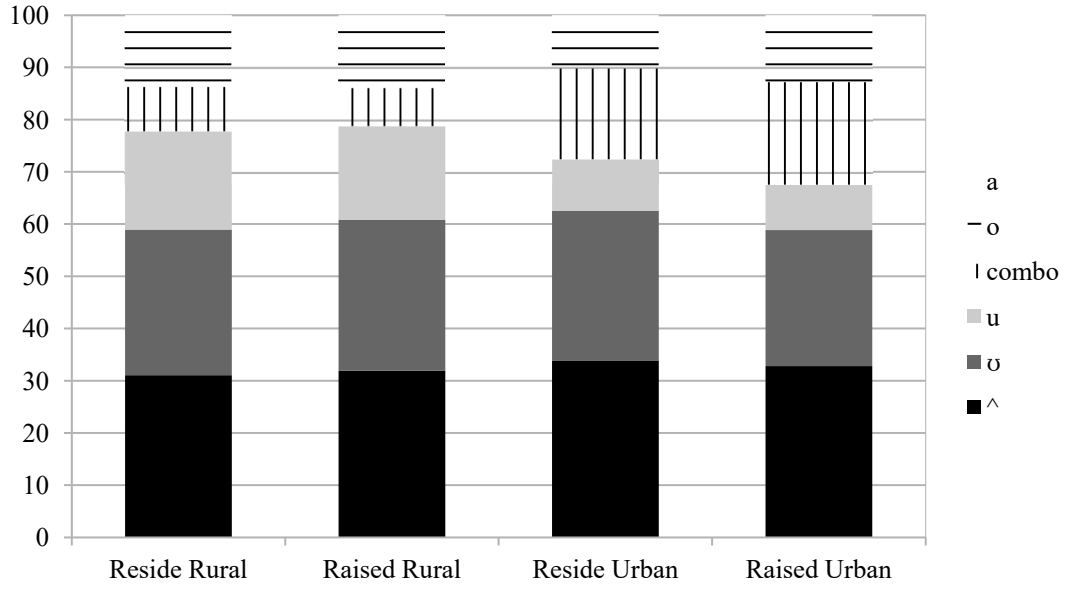

Figure 9. Vowel matches made to the word *full*

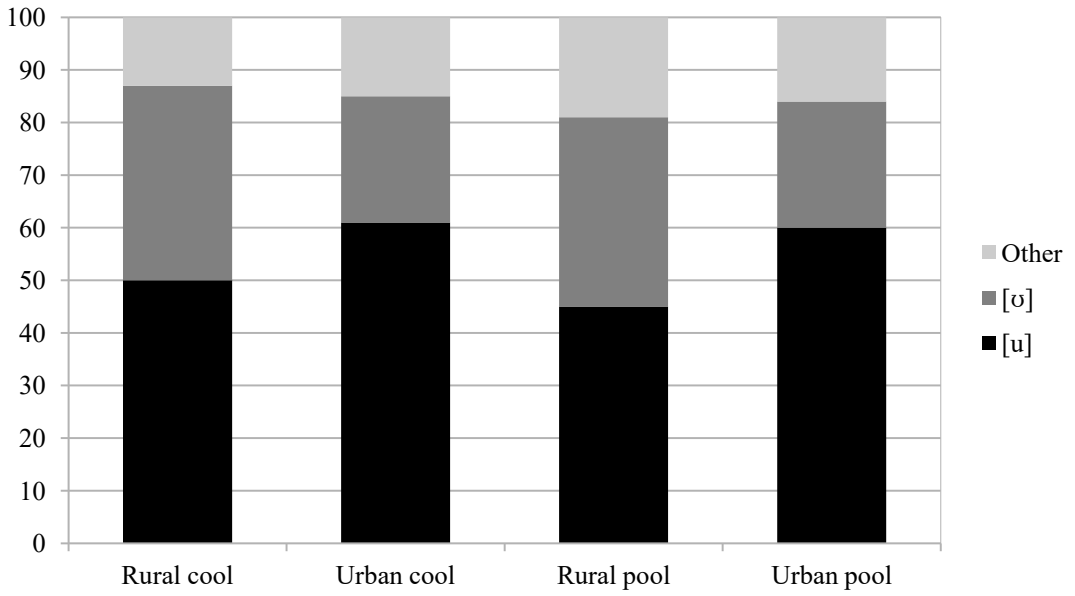

Figure 10. Vowel matches made to *cool* and *pool*

Cool and *pool* pattern together quite closely (Figure 10). In both words, Utahns raised in urban areas matched them with the test response *Luke* more than rural Utahns, who favored [ʊ] by matching their pronunciations with *look* (Tables 3 and 4). The merger between [i] and [ɪ] is apparent in the test word *meal* where 23% of the responses by rural Utahn residents matched the lax vowel in comparison to 14% of urban residents (Table 6).

5.8 Intrusive [t]: across

A highly stigmatized feature of Utah English is the appearance of intrusive [t] in words such as *also, censor,* and *answer.* While preference for transitional [t] in these words did not reach significance, word-final epenthesis on across did. Twelve percent of urban residents recognized their pronunciation of [t] compared to 32% of rural Utah residents (Table 5).

6. Utah Results of the Harvard Dialect Survey

Another source of data that can be used to find Utah traits that differ between rural and urban areas is the Harvard Dialect Survey (Vaux & Golder, 2003). This survey was posted online and garnered over 47,000 responses to 122 questions about pronuncation, vocabulary, and syntactic usage. Of the respondents, 322 were from Utah. However, the Utah results must be considered with caution since 301 of the participants come from urban areas and only 21 from rural areas.

With this in mind, three predictors were found to be significant (Table 7). The pronunciation of the grapheme 's' in the surname *Presley* is [s] for 90% of the rural Utahns and only 74% for those from urban areas. When asked what the underground part of a building is called 43% of rural Utahns distinguised between a *basement* that is finished, and a *cellar* that is unfinished. In contrast, only 16% of urban Utahns made the same distinction. A circular intersection was referred to as a *roundabout* by 58% of urban Utahns and only by 29% of rural Utahns, many of whom preferred *traffic circle* instead.

Table 7. Logistic regression results of the Harvard Dialect Survey.

Predictor	Estimate	Z	p	Odds ratio	% A Urban	% A Rural	%B Urban	%B Rural
Presley: A [z], B [s]	2.1	1.97	.049	8.33	21	5	74	90
Underground level: A *basement,* B *Basement* is finished, *cellar* is not	2.01	3.18	< .001	745	60	29	16	43
Circular intersection: A *roundabout,* B *traffic circle*	2.14	3.32	< .001	8.61	58	29	13	43
Circular intersection: A no word, B *traffic circle*	3.20	2.50	.013	24.53	11	5	13	43

McFadden's $R^2 = .232$

7. Discussion

The division of the state into rural and urban dialects was noted several decades ago by Lillie (1993). In an analysis of the Utah English Survey the responses of participants who were raised in or reside in the Wasatch Front often ranked closer to those participants who were either not raised in Utah or did not live in the state when they took the survey. In other words, urban Utah speech is less Utah-like than the speech of rural Utahns. Many of the stereotypes of Utah dialect are actually more prevalent in the rural areas, specifically vowel mergers before [ɬ] (e.g. *meal, full, pool*), *Laura* with [ɑ], intrusive [t] at the end of *across*, and *measure* with [eɪ].

The variation between *pop* and *soda* emerges as the most telling difference between urban and rural dialects. The encroachment of *soda* on the more traditional *pop* is attributed to a spread of that term from states where *soda* dominates, more specifically to immigrants from California, and to a lesser extent Arizona. In recent decades Californians have comprised the largest group of immigrants to Utah (Gregory 2017; Figure 11), and may have brought their *soda* with them.

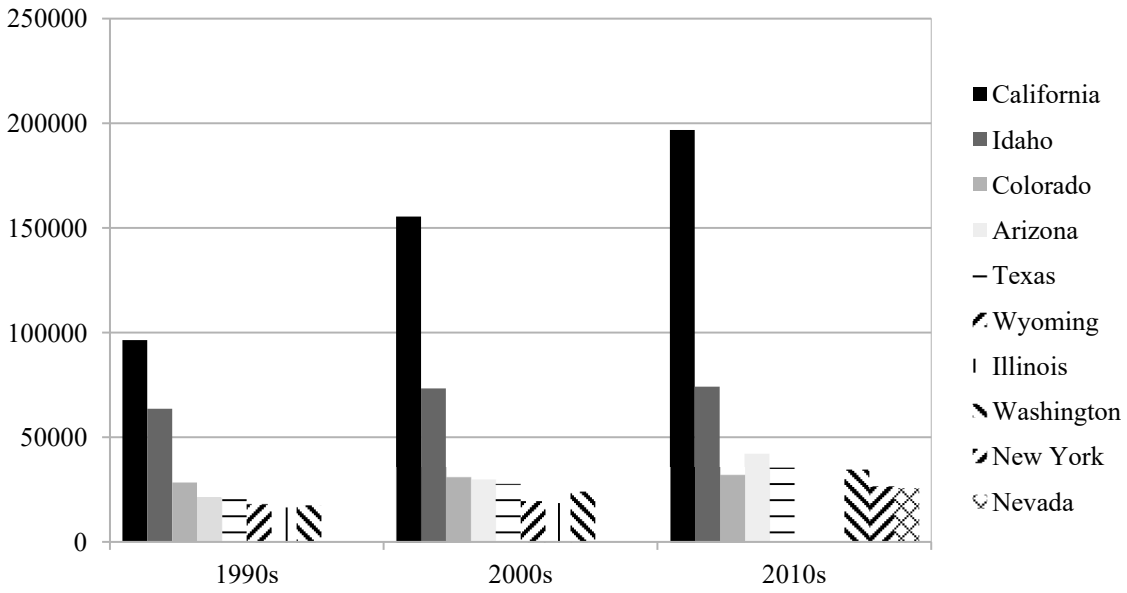

Figure 11. The Eight States that Provided the Most US Born Immigrants to Utah.

The population of Utah has doubled in the last 40 years from about 1.5 million in 1980 to over 3 million. While Utah does have a high birth rate, much of the population boom is due to out-of-state immigrants. If immigrants gravitated towards the urban areas of the state a case could be made that their influence may have aided in distinguishing the rural and urban dialects. In order to test this, the proportion of Utahn residents in each county in 2019 who were born in the state was obtained (Stacker, 2022) and submitted to a Welch's t-test to compare the counties in the rural and urban dialect areas. There were no significant differences between them ($t(26.3) = .779$, $p = .443$) which discredits the uneven immigration hypothesis.

The linguistic division of the counties into rural and urban dialects patterns quite closely with a number of demographic factors as well. A Welch's ANOVA with data from the 2010 census using the two dialect areas as the dependent variable reveals that counties belonging to the urban dialect have a higher population, a lower poverty rate, more residents under 18, and fewer over 65 (Table 8).

Table 8. ANOVA results of demographics by county belonging to two dialects areas.

	F	df1	df2	*p*	Mean Urban	SD Urban	Mean Rural	SD Rural
Population	5.83	1	10.0	.036	274290	357179	14134	11905
Poverty rate	4.53	1	26.7	.043	8.51	2.83	11.31	4.25
% of adults with bachelors degree	4.61	1	24.7	.042	30.81	8.18	23.45	10.10
% of population under 18	5.30	1	27.0	.029	30.59	2.66	27.57	4.40
% of population over 65	12.84	1	26.4	.001	11.79	3.59	17.61	5.13

The fact that the urban area is comprised of more young residents and fewer older ones is telling. In the Utah Dialect Survey, a number of features were more prevalent among older Utahns, namely the use of *pop,* pronouncing *measure* with [eɪ], accepting for + adjective interjections, pronouncing *Laura* and *Lauren* with [ɑ], and pronouncing [ɬ] in words such as *stalk*. It is these same dialect features that are more common in rural areas. This suggests that the dialect distinction may, in part, be attributed to age differences. In other words, older speakers maintain older dialectal characteristics, and they also happen to comprise a larger portion of the rural population. However, the relationship between age and rural traits does not hold up everywhere. Older participants in the Utah English Survey retained [u] in words like *pool* and *cool* to a greater degree than younger speakers, yet the [u] pronunciation was more common along the urban Wasatch Front.

8. Conclusions

The purpose of this paper has been to examine previous suggestions about dialect boundaries in Utah. Neither Lillie's (1998) nor Eddington's (2023) are based on empirical data. For this reason, a statistical analysis of the Utah English Survey was used to draw dialect boundaries in the state. The two dialect regions that emerged correspond to the rural and urban regions of the state. Arriving at these dialect regions by computational means has a number of advantages. First, it is based on the results of the entire Utah English Survey, rather than by hand picking particular features. Second, it is objective in the sense that features were not chosen or excluded in advance. Of course, the features included in the survey itself influenced the outcome. However, they include most of the phenomena that have been discussed in the literature on Utah dialect. Third, the two dialect boundaries aligns quite well with other social variables such as population, poverty level, and age.

Of course, any dialect division that only factors geography into the equation ignores important differences that exist within a single dialect area such as gender, age, education, and religion. These differences were reported in the results of Utah English Survey. However, the resulting dialect areas may serve as a guide for future linguistic study of the state.

References

Abadi, M. (2018). 'Soda,' 'Pop,' or 'Coke': More than 400,000 Americans Weighed in, and a Map of Their Answers Is Exactly What You'd Expect, *Business Insider,* October 6. Retrieved from https://www.businessinsider.in/soda-pop-or-coke-more-than-400000-americans-weighed-in-and-a-map-of-their-answers-is-exactly-what-youd-expect/articleshow/66074186.cms

Baker-Smemoe, Wendy, & David, B. (2015). Linguistic Behavior and Religious Activity. *Language & Communication, 42*, 116-124. https://doi.org/10.1016/j.langcom.2014.12.004

Baker, Wendy, David, E., & Lyndsey, N. (2009). Dialect Identification: The Effects of Region of Origin and Amount of Experience. *American Speech, 84*(1), 48-71. https://doi.org/10.1215/00031283-2009-004

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences, 10*(2-3), 191-203. https://doi.org/10.1016/0098-3004(84)90020-7

Blankenship, B. (1992). What TIMIT Can Tell Us About Epenthesis. *UCLA Working Papers in Phonetics*, *81*, 17-25.

Bowie, D. (2003). Early Development of the Card-Cord Merger in Utah. *American Speech, 78*(1), 31-51.

DeBry, R. J., & Associates. (2109). These 11 Phrases Are a Dead Giveaway That You're from Utah. *Deseret News.* June 17. Retrieved from https://www.deseret.com/2019/6/21/20794577/these-11-phrases-are-a-dead-giveaway-that-you-re-from-utah.

DiPaolo, M. (1993). Propredicate Do in the English of the Intermountain West. *American Speech, 68*(4), 339-356. https://doi.org/10.2307/455771

Eddington, David, & Ellingson (2023 forthcoming). *Utah English.* Salt Lake City UT: University of Utah Press.

Eddington, David, & Matthew, S. (2012). Where Are the Moun[ʔə]ns in Utah? *American Speech, 87*, 336-349. https://doi.org/10.1215/00031283-1958345

Graham, J. (2006). 'Oh For' as a Scandinavian-influenced Linguistic Feature of Minnesota and Utah. (2006). Paper presented at the 60th Annual RMMLA Convention, Tucson, AZ, October 12-14.

Grieve, J., Speelman, D., & Geeraerts, D. (2011). A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, *23*(2), 193-221. https://doi.org/10.1017/S095439451100007X

Holmes, G., Pfahringer, B., Kirkby, R., Frank, E., & Hall, M. (2002). Multiclass alternating decision trees. In *European Conference on Machine Learning,*161-172. Berlin, Heidelberg: Springer.

The Jamovi Project. (2020). *Jamovi*. V. 1.2. Linux/Mac/Windows. 2020. Retrieved from https://www.jamovi.org/about.html

JASP Team. (2022). JASP. V. 0.16.01. Retrieved from https://jasp-stats.org/

Lillie, D. D. (1998). *The Utah Dialect Survey.* Master's thesis, Brigham Young University.

Lillie, D. D. Utah Dialects. Retrieved from http://www.geocities.ws/diandyl/udp/dialects.html

Nerbonne, J. (2009). Data-driven dialectology. *Language and Linguistics Compass, 3*(1), 175-198.

https://doi.org/10.1111/j.1749-818X.2008.00114.x

Pardoe, T. E. (1935). Some Studies of Rocky Mountain Dialects. *Quarterly Journal of Speech, 21*(3), 348-355. https://doi.org/10.1080/00335633509380117

R Core Team. R. (2020). A Language and Environment for Statistical Computing. V. 3.6. Linux/Mac/Windows, 2020.

Savage, D. M. (2014). *How We Feel about How We Talk: A Language Attitude Survey of Utah English.* Master's thesis, Brigham Young University.

Seol, H. (2020). *SnowCluster: Cluster Analysis.* [jamovi module]. Retrieved from https://github.com.hyunsooseol/snowCluster

Séguy J. (1971). La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35, 335-357. http://doi.org/10.5169/seals-399508

Wikipedia. (2021). Wasatch Front. Last modified October 6. Retrieved from https://en.wikipedia.org/wiki/Wasatch_Front

Wikitravel. (2022). Wasatch Front. Last modified February 7. Retrieved from https://wikitravel.org/en/Wasatch_Front

World Population Review. (2020). Accessed December 30. https://worldpopulationreview.com/us-counties/states/ut

Yoo, I. W., & Blankenship, B. (2003). Duration of epenthetic [t] in polysyllabic American English words. *Journal of the International Phonetic Association*, *33(*2), 153-164. https://doi.org/10.1017/S0025100303001269