

Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Yale Working Papers in Grammatical Diversity

Department of Linguistics

10-2021

Variations on Defining a Dialect Region

Sasha Lioutikova

Follow this and additional works at: <https://elischolar.library.yale.edu/ygdp>



Part of the [Syntax Commons](#)

Variations on Defining a Dialect Region

Sasha Lioutikova^{1*}

Abstract

This paper explores various methods of defining a dialect region, using syntactic constructions characteristic of the American South. We consider three important characteristics of such a method: place, feature, and measure. "Place" details how data points should be grouped, "feature" details which data points should be used, and "measure" details how the data should be analyzed. While we may not claim one method as superior to any other, this research can provide insight into how different techniques of characterizing geographic variation may affect the resulting dialect regions.

Keywords

dialect regions — dialectology — Southern American English

¹ Yale University

*Corresponding author: sasha.lioutikova@yale.edu

1. Introduction

In this paper, we will explore how different methods of determining a dialect region yield different results, and the extent to which these resulting regions correspond to previously established dialect regions. The data used in this exploration are acceptability judgements from the Yale Grammatical Diversity Project (YGDP) surveys (Wood et al. 2020b). We will be exploring different ways of setting three parameters for using the YGDP data to generalize across regions:

- (a) What counts as a place?
- (b) What data will be considered for that place?
- (c) How will that data be measured?

For (a), we will compare hot spot regions (Wood 2019), hand-drawn isoglosses, and tessellations. Each of these methods will be discussed in more detail in what follows. For (b), we will compare two data sets: a set of two sentences representative of each construction (for a predetermined set of constructions) and a set of half of all sentences for each type of construction. For (c), we will compare using average, minimum, and maximum measurements within a region.

These methods will result in different characterizations of geographical dialect regions. We will compare the resulting regions of each methodology to the regions proposed in the *Atlas of North American English* (ANAE) (Labov et al. 2006) to see which line up most closely with

those regions. While the present project and the ANAE used different types of data (syntactic acceptability judgment data and phonetic/phonological data, respectively), it has previously been noticed that many of the major dialect regions can be replicated using different methods and different kinds of data. For example, Labov et al. (2006) note that the results of the ANAE are similar in many ways to the results in Carver’s (1987) study, which was based almost entirely on lexical variation. In the present study, the methods that yield results most closely matching the ANAE are (a) the use of hot spot regions, (b) using half of the sentences for each construction, and (c) taking the average of the judgments. The results suggest that these methods could be considered sound practices as starting points for characterizing a dialect region using acceptability judgment data.

2. Place

When building dialect regions from the bottom up, it is not immediately obvious what should count as a “place.” Each survey participant is linked to a place—the place that they reported as their primary childhood residence (Wood et al. 2020b). However, we cannot simply take these coordinates as a “place” in the context of this exploration; dialect regions are determined not by only one individual’s linguistic habits, but by the linguistic habits of groups of speakers. Thus, in this context, it is necessary to move from an individual to a grouping of individuals in our definition of “place.” To do so, we must decide how to group participants together into a region. In this paper, we explore three different ways of defining a place for dialect region analysis: hot spot regions defined on Voronoi polygons, hand-drawn isoglosses, and tessellations.

2.1 Data

In this paper, we focus on the dialect geography of the South. We therefore chose sentences that we know to be characteristic of the South, specifically, Personal Dative and Presentative Dative sentences whose Hot Spots maps show clear regionality in the South (Wood et al. 2015, 2020a; Wood and Zanuttini 2018). The full set of sentences used for this section is presented in Table 1, along with their sentence identifiers.

For all of these sentences, we created a point layer that contained only the respondents that judged that sentence using the Select tool in ArcGIS Pro,¹ with the full YGDP data set point data as the Input Feature. This is done for all of the methodologies in this project.

2.2 General Methodology

In order to facilitate a comparison of the different methodologies, we first define a base methodology that all of the approaches to “place” will use. For each sentence, we created regions of significance using its acceptability score along with the specified method of defining a place. Those regions were converted to raster layers in ArcGIS Pro and given a value of 1. We then layered all of the sentence regions using Cell Statistics; the Input rasters were all of the significant sentence regions, and the Overlay statistic was Sum (so that areas with more overlapping sentence regions will have a higher value, reflecting the number of overlaps). We then smoothed all of the sentence regions using Focal Statistics to give a clearer visual representation of the resulting dialect region (cf. Sibling et al. 2012; Stoeckle 2014), and to reduce the arbitrariness of specific border edges

¹Title caps will be used when discussing specific tools in ArcGIS Pro. Note that many of the basic technical aspects of GIS processing will not be explained in this paper. The reader is invited to consult with the ArcGIS Pro website (<https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>) for more information.

Sentence Number	Sentence
F1002	“Here’s you a piece of pizza.”
F1007	“Where’s us some ripe blueberries?”
F1009	“Where are me some country boys?”
F1067	“Where’s me a screwdrvier?”
F1068.1	“Okay, now where’s you a pillowcase?”
F1069	“Where’s us a good place to eat around here?”
F1070	“Here’s me a good pair of jeans!”
F1070.1	“Here’s me a good pair of jeans.”
F1071	“Here’s us a gas station — pull over!”
F1098	“Here’s him a nice cup of coffee.”
F1099	“Here’s John a glass of iced tea.”
F1116	“Here’s you some money.”
F1117	“There’s you a piece of pizza.”
F1118.1	“Here are you some books.”
F1119	“Here is you a new bunny.”
F1120	“Here is you some camping ideas.”
F1121	“Here comes you a bus.”
F1122	“Here’s you some fun ideas.”
F1123	“Now there’s me a new Easter dress.”
F1124	“Now there’s us a story.”
F1125	“Now there’s us some easy money.”
F1179	“I’m gonna go and play with me a cat.”
F1180	“I’d put me a marble or two in my pocket. ”

Table 1. Dative Sentences

(effectively making such edges gradient rather than discrete). The Input raster for this tool was the layered raster resulting from Cell Statistics, and the Neighborhood was a Circle with a radius of 20 cells.

2.3 Hot Spots Region

Following the methodology outlined in Zanuttini et al. (2018), Wood (2019), and Wood et al. (2020b), we first created Voronoi polygons to use as input for the Hot Spots Analysis. To do so, we used the tool Create Thiessen Polygons. The Input was all data (as a points layer) and the Output fields was ALL. We clipped the Voronoi polygon layer to the shape of the United States.

We then created Hot Spots for each sentence. To do so, we used the tool Hot Spots Analysis, which uses the Getis-Ord G_i^* statistic (Ord and Getis 1995). The Input feature was the (clipped) Voronoi layer, the Input field was the sentence for which we were creating Hot Spots (ex. F1002), and the Distance band was 450,000 meters. This distance band was the same for all sentences; it gave Hot Spot regions that are not too generous in area, but still mostly continuous.

For analysis purposes, we then used the tool Polygon to Raster to convert the Hot Spots polygons to rasters. The Input feature was the Hot Spot layer, the Value field was the GiZScore, and the Cellsize was 10,000 units. We then used the tool Reclassify to identify areas that accepted that sentence statistically significantly more than the rest of the sample (i.e. areas of significance).

The Input raster was the Hot Spot raster (from the previous step) and the Reclass field was Value (GiZScore, from the previous step). All GiZScores greater than or equal to 2 were reclassified to 1, and all GiZScores less than 2 were reclassified to 0. In other words, we identify all areas for which the GiZScore is greater than or equal to 2 as significant. We do this because areas whose acceptability judgements are more than 2 standard deviations away from the mean are likely not explained by random variation; the cutoff of 2 for GiZScores corresponds to a p value of around 0.0455, less than the standard statistical cutoff of 0.05.

We then stacked and smoothed these areas using Cell Statistics and Focal Statistics, as described in the general methodology. The results are shown in Figure 1, where the lighter red shows areas where hot spot regions for 12–18 sentences overlap, and the darker red shows areas where hot spot regions for 18–23 sentences overlap.²

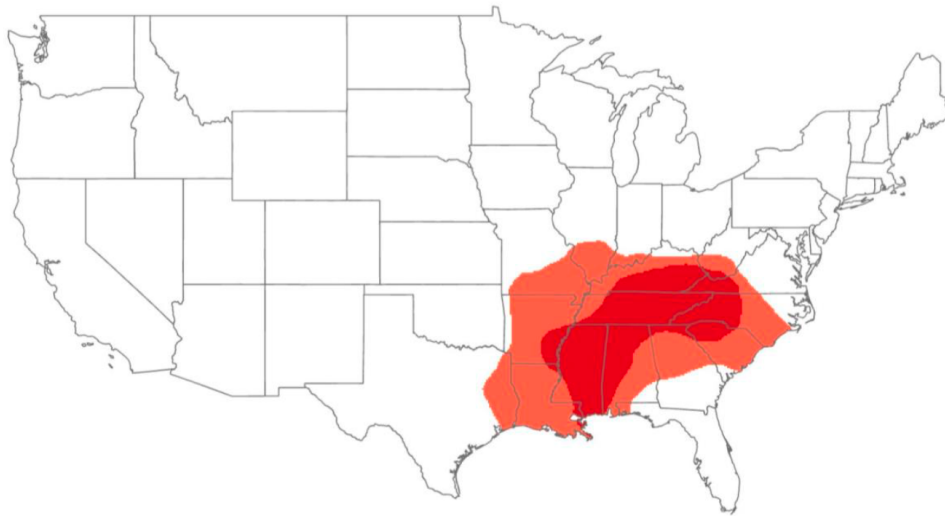


Figure 1. Hot Spots Dialect Region: Lighter Red: 12–18, Darker Red: 18–23

2.4 Hand-Drawn Isoglosses

In order to draw our own polygons in ArcGIS Pro, we first needed to create an empty polygon layer for each sentence. To do so, we used the tool Create Feature Class. The Geometry type was Polygon and the Coordinate System was Current Map [Map].

For each sentence, we changed the symbology of the sentence’s points layer—the layer that displays the participants—to easily see whether a participant accepted or rejected the sentence and where that participant is located. We chose unique values, and set Field 1 (which determines what data to present) to the sentence for which the polygon is being drawn (ex. F1002). We grouped values 1 and 2 and changed the symbol to a black dot, grouped values 4 and 5 and changed the symbol to a green dot, and removed the value 3. To draw the significant polygons with the points as a reference, we used the Create Features tool (Edit > Create) and selected Freehand to hand-draw a reasonable significant region. A map of all of the polygon layers is shown in Figure 2.

²Due to the averaging function of the Focal Statistics tool, this is a gradient value, so 18–23 is really something like 18.00001 to 23.

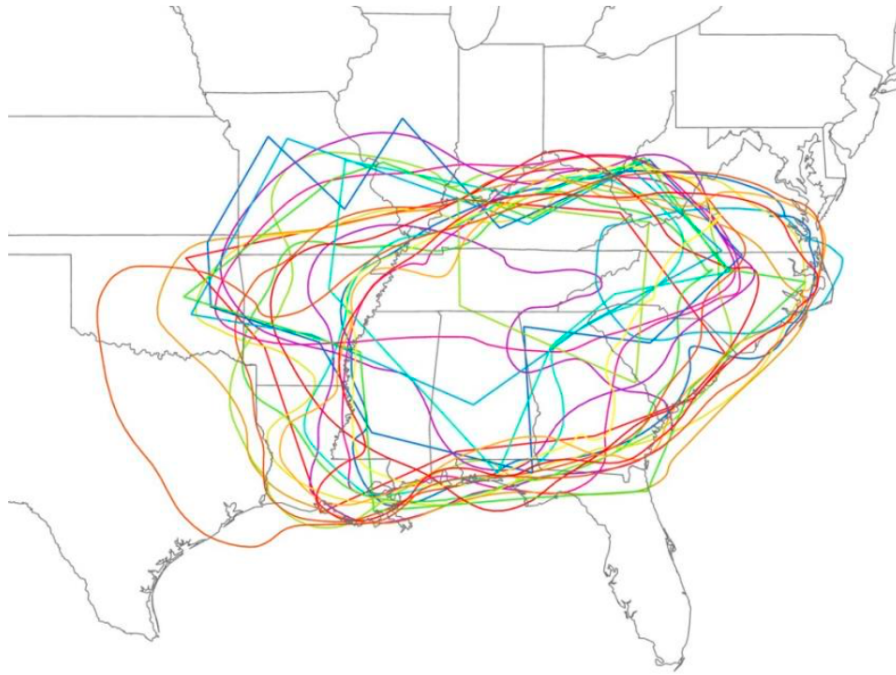


Figure 2. All Hand-Drawn Polygons

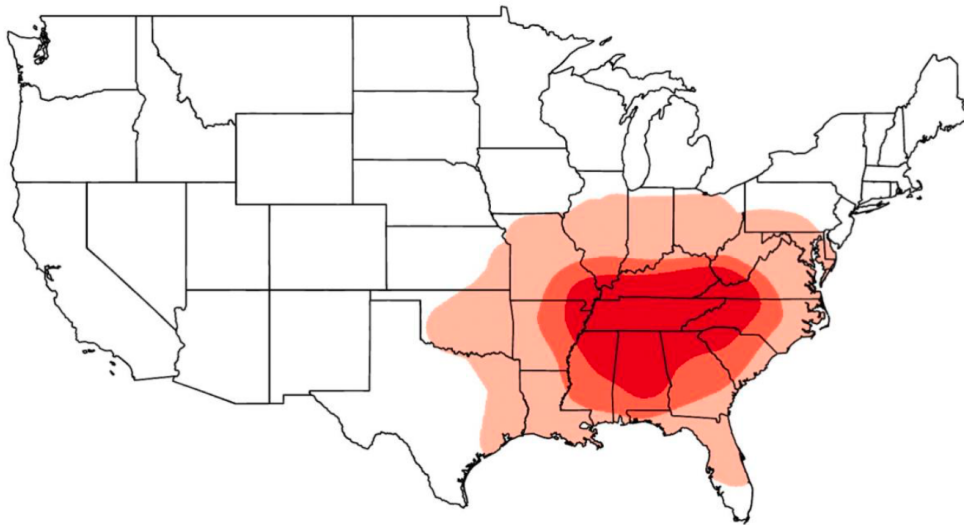


Figure 3. Hand-Drawn Isoglosses Dialect Region: Lightest Red: 8–12, Medium Red: 12–18, Darkest Red: 18–23

After all the polygons were made, we converted them to raster using the tool Polygon to Raster (again, for analysis purposes). The Input feature was the hand-drawn polygon, and the Value field was OBJECTID (automatically 1, so no Reclassification required like in the Hot Spots methodology). We then stacked and smoothed these areas using Cell Statistics and Focal Statistics, as described in the general methodology. The resulting map, with the same symbology as the Hot Spot map, is shown in Figure 3.

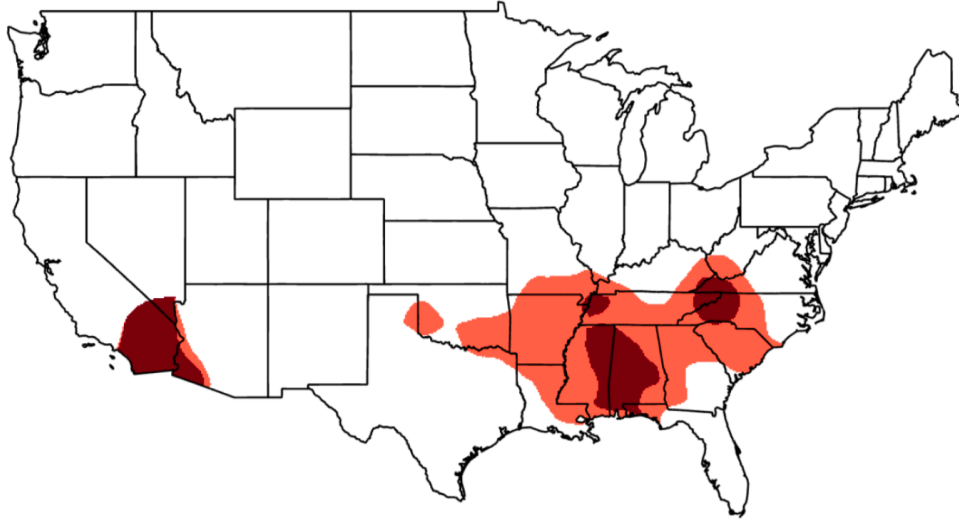


Figure 4. Tessellations Dialect Region. Lighter Red: 9–14, Darker Red: 14–18 (See section 2.6 for discussion of the hot spot in California.)

2.5 Tessellations

We first created a tessellation to split up the country into equally-sized polygons using the tool Generate Tessellation (see Figure 8). The Shape Type was Hexagon and the Size was 40,000 square kilometers (the default settings). We then performed a Spatial Join for each sentence, where the Target Feature was the tessellation, the Join Feature was the sentence point data, and the Merge Rule was Mean (i.e. the value of the hexagon would be set to the average judgements of the points within it). We then converted the all of the tessellation layers to raster (Polygon to Raster, just like the previous two methods). The Input feature was the sentence’s tessellation and the Value field was the sentence. We then used the tool Reclassify to identify areas of significance. The Input raster was the tessellation raster (from the previous step) and the Reclass field was Value (acceptability judgement, from the previous step). Judgements of 4 or 5 were reclassified to 1, and all other judgements were reclassified to NODATA. Thus, hexagons whose average judgement was 4 or 5 were considered significant areas. We then stacked and smoothed these areas using Cell Statistics and Focal Statistics, as described in the general methodology. The resulting map, while having different values for its symbology, still reflects the same proportion of accepted regions within each category (i.e. lighter red accepts 50-75% of overlapped sentences, darker red accepts 75-100% of overlapped sentences). This map is shown in Figure 4.

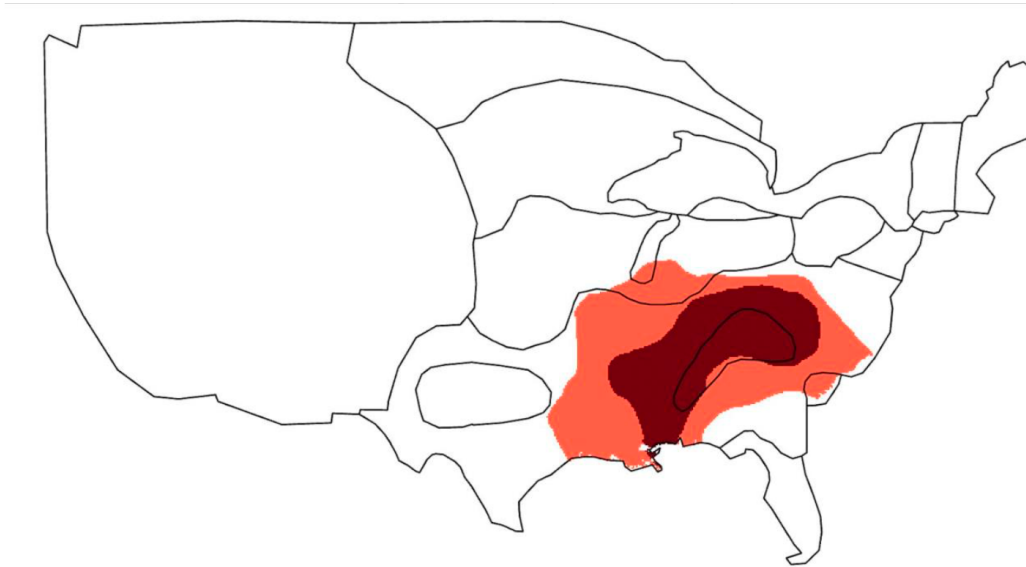


Figure 5. Hot Spots Dialect Region w/ ANAE Regions. Lighter Red: 12–18, Darker Red: 18–23

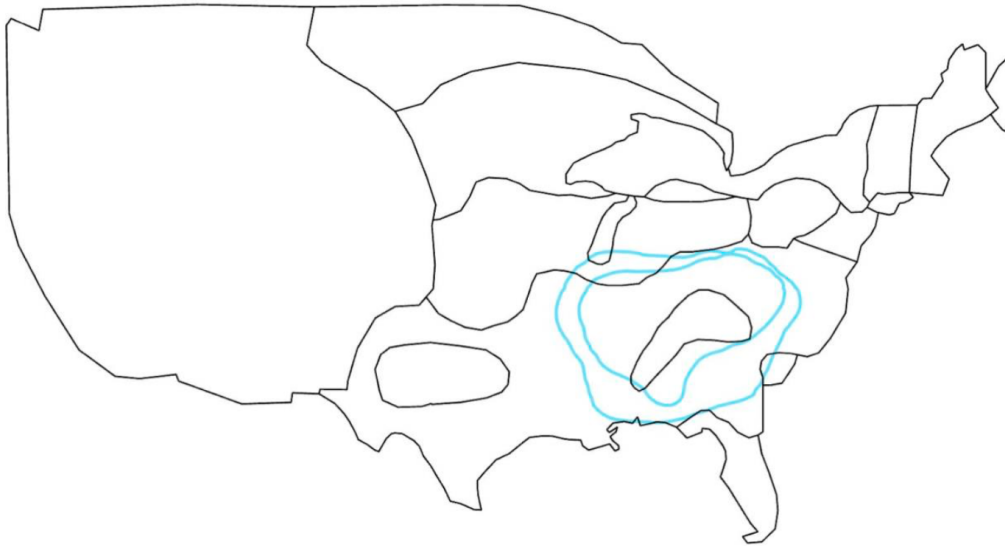


Figure 6. Hand-Drawn Isoglosses Dialect Region w/ ANAE Regions. Inner region: 12–18, Outer Region: 18–23

2.6 Analysis

In Figures 5–7, we present the regions established from each of these methods on top of the regions proposed in the ANAE. It is clear that the hot spots method, shown in Figure 5, provided the closest match. It best represents a reliably specific, continuous region. Moreover, it even seems to pick out something quite close to the Inner South within the general South region.

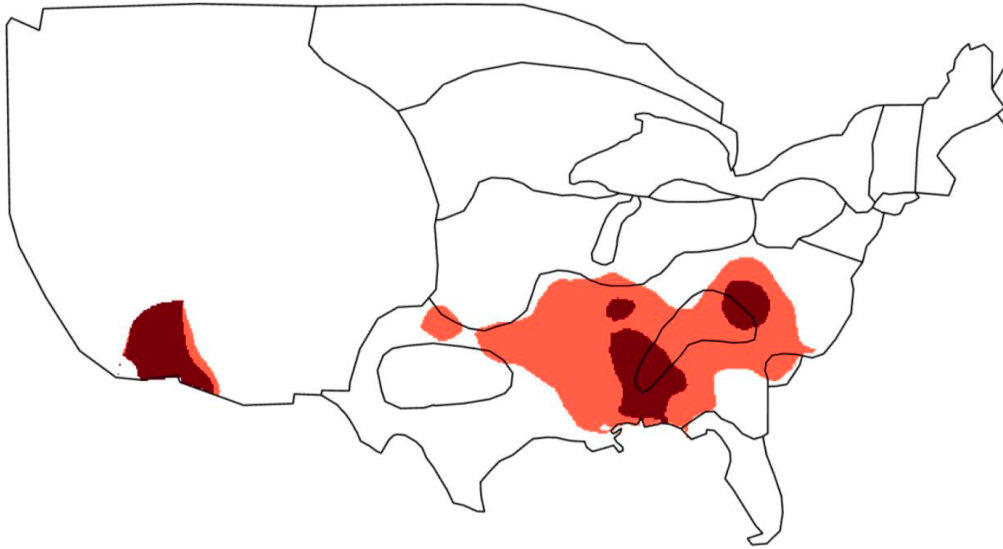


Figure 7. Tessellations Dialect Region w/ ANAE Regions. Lighter Red: 9–14, Darker Red: 14–18

The Hot Spot methodology better takes into account clusters of high judgements than the tessellation methodology. The tessellation methodology treats a polygon with many high acceptability judgements the same as a polygon with a few high acceptability judgements, given that their average is the same. In this way, the tessellation methodology loses important information on the frequency of acceptability judgements of a sentence within a region. For example, the noise in the tessellation maps (Map 4, Map 7) in lower California is the result of only two people who happen to accept characteristically Southern sentences in one polygon, shown in Figure 8 below.

Taking into account the dialect region maps and the analysis concerning the methodologies themselves, we suggest that the Hot Spot overlap methodology is the best out of the three for identifying a dialect region from the bottom up.

3. Feature

In the analysis of a dialect region, the way a feature is defined potentially affects the shape of the resulting regions. We will explore two different methods of defining a feature for dialect region analysis: choosing two sentences per construction, and choosing half of all available sentences per construction. That is, if there are eight sentences total for one construction, four would be chosen for analysis.

3.1 Data

The data that we have used for this analysis comes from the YGDP dataset described in Wood et al. (2020b). Because we are looking specifically at the South, we chose five constructions that we know to be characteristic of the South: personal datives, dative presentatives, extended benefactives, *fixin' to*, and split subjects. Within one construction, all sentences were given a number, and a number of sentences were chosen for analysis using a random number generator. The selected sentences for both methods are shown in Tables 2 and 3.

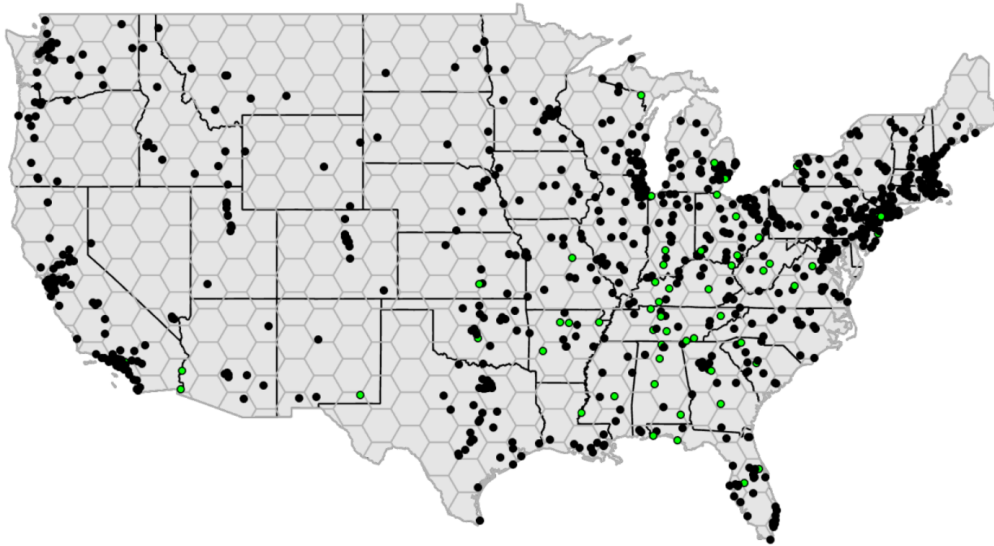


Figure 8. Tessellations Dialect Region w/ Point Data. F1100 (“I hunted the hills over for you a squirrel.”) Black: 1–2, Green: 4–5

3.2 General Methodology

The base procedure for creating a dialect region for this aspect of dialect region analysis is based off of the Hot Spot methodology discussed in the “place” methodology. The procedure was followed up to the Reclassify step (Create Thiessen Polygons, Hot Spot Analysis, Polygon to Raster, Reclassify) for each sentence. This created significant regions for each sentence. Then, for each construction, significant regions were made by layering all sentences belonging to that construction (Cell Statistics) and classifying a significant region (Reclassify). Then, just like the Hot Spots methodology, the five construction regions were layered (Cell Statistics) and smoothed (Focal Statistics).

3.3 Analysis

The resulting maps for both of the “feature” methodologies are shown without overlaid ANAE regions in Figures 9 and 10 and with overlaid ANAE regions in Figures 11 and 12. Note that they have the same symbology. The two methodologies create extremely similar maps – so similar that it is difficult to determine which method is more sound using solely the maps. In terms of the methodologies themselves, the procedure that uses half of all sentences per construction is more reliable, as there is less information lost. Sentences belonging to one construction vary in markedness depending on the particular structure of the sentence (e.g. *here* vs. *there* vs. *where* for a dative presentative construction). Randomly choosing only two sentences runs the risk of misrepresenting the true regional distribution of the general construction. If we choose half of the sentences, there is a greater chance that the sentences we choose will be characteristic of the regional distribution. As the dataset becomes larger, however, it may become far too tedious to analyze half of all of the sentences per construction. Perhaps the constructions may be classified into more specific groups, and a few sentences may be chosen from each group to create a more

Sentence Number	Sentence
Personal Datives	
F1095	“I need me some black jeans.”
F1096	“She has her a new boyfriend.”
F1179	“I’m gonna go and play with me a cat.”
F1180	“I’d put me a marble or two in my pocket.”
Dative Presentatives	
F1009	“Where are me some country boys?!”
F1067	“Where’s me a screwdriver?”
F1068	“Where’s you a quiet place to study?”
F1070	“Here’s me a good pair of jeans!”
F1098	“Here’s him a nice cup of coffee.”
F1116	“Here’s you some money.”
F1117	“There’s you a piece of pizza.”
F1121	“Here comes you a bus.”
F1122	“Here’s you some fun ideas.”
F1123	“Now there’s me a new Easter dress.”
F1125	“Now there’s us some easy money.”
Extended Benefactives	
F1100	“I hunted the hills over for you a squirrel.”
F1101	“We are looking for him a new home.”
F1178	“I’ll be right back with you some tea, okay?”
Fixin’ To	
F1208	“The boss is fixin’ to buy a bunch of new desks this year.”
F1209	“I’m fixin’ to get these floors redone.”
F1210	“My brother’s fixin’ to bring his family for a visit.”
Split Subjects	
F1038	“But those people won’t any of them fight fair.”
F1073	“They don’t any of them need any advice anymore, because they’ve been going to school for two years.”
F1075	“We don’t any of us politicians tell the truth all the time.”
F1077	“They won’t any of them tell the truth.”

Table 2. Sentences for Feature Analysis (Half Sentences)

representative subset of the construction. With the current YDGP dataset, though, the method of choosing half of all sentences per construction appears to be reasonable and more sound.

Sentence Number	Sentence
Personal Datives	
F1095	“I need me some black jeans.”
F1097	“He needs him that big truck over there.”
Dative Presentatives	
F1098	“Here’s him a nice cup of coffee.”
F1125	“Now there’s us some easy money.”
Extended Benefactives	
F1100	“I hunted the hills over for you a squirrel.”
F1102	“I have him a new book.”
Fixin’ To	
F1206	“She’s fixin’ to mow the lawn.”
F1208	“The boss is fixin’ to buy a bunch of new desks this year.”
Split Subjects	
F1038	“But those people won’t any of them fight fair.”
F1077	“They won’t any of them tell the truth.”

Table 3. Sentences for Feature Analysis (2 Sentences)

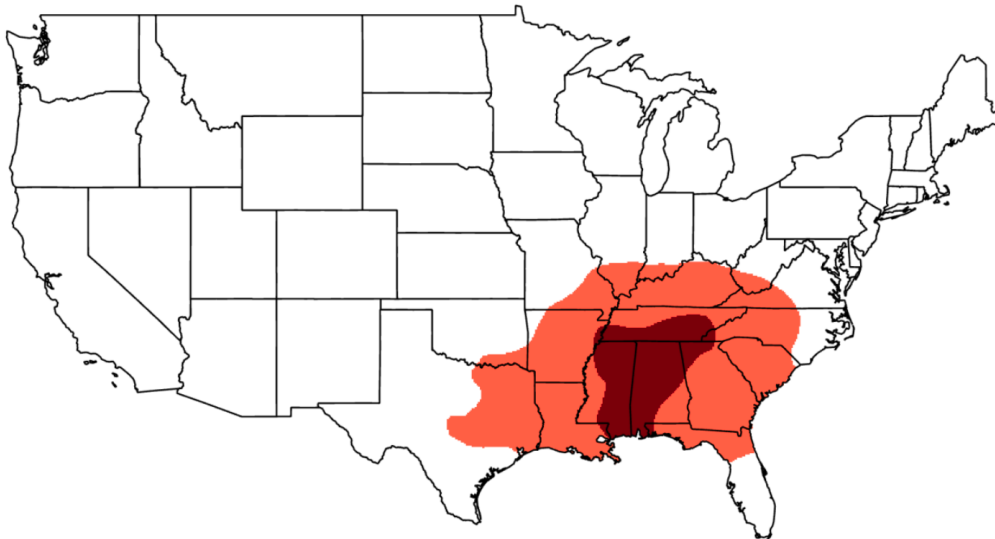


Figure 9. Two Sentences Dialect Region. Lighter Red: 3–4.5, Darker Red: 4.5–5

4. Measure

In analyzing regional dialect variation, different ways of measuring or analyzing the data may affect the resulting dialect regions. In this section, we will explore three different definitions of measure: mean, minimum, and maximum.

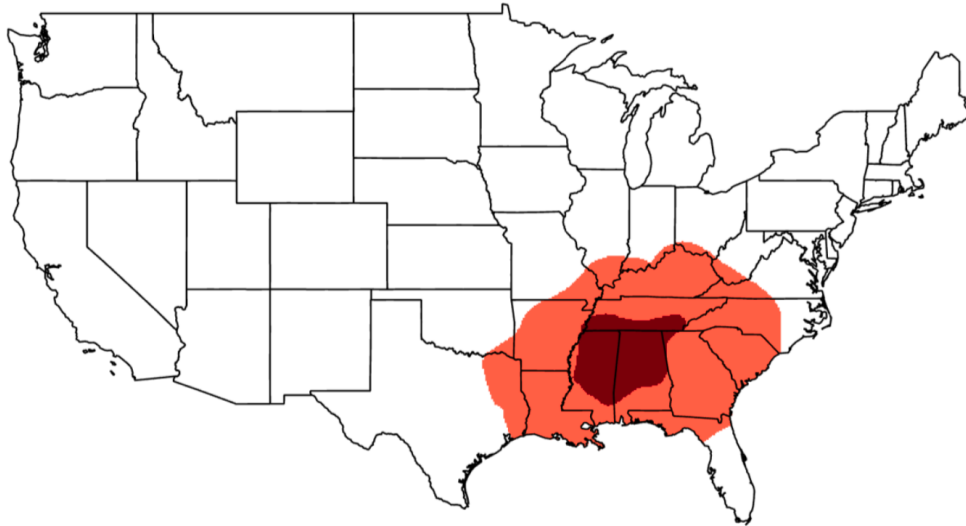


Figure 10. Half of All Sentences Dialect Region. Lighter Red: 3–4.5, Darker Red: 4.5–5

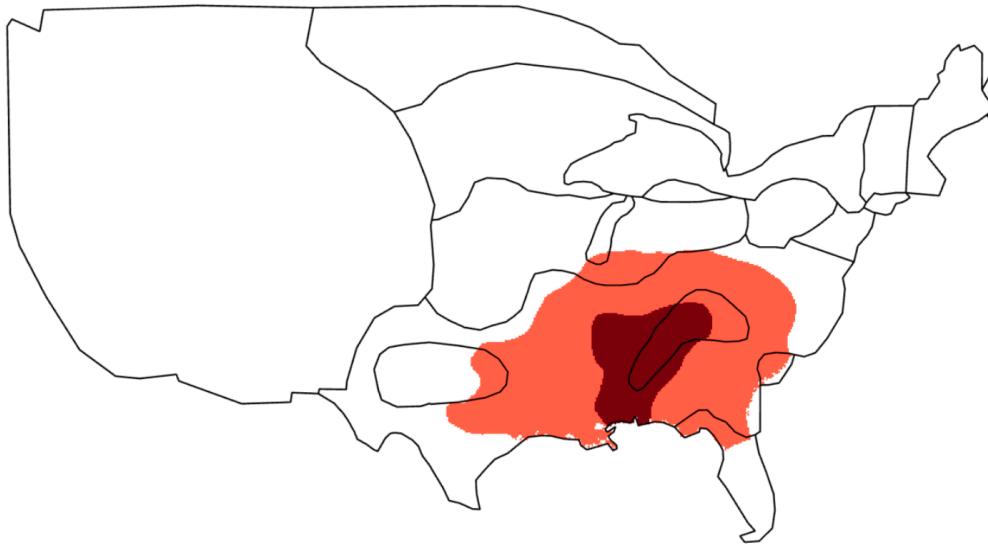


Figure 11. Two Sentences Dialect Region w/ ANAE Regions. Lighter Red: 3–4.5, Darker Red: 4.5–5

4.1 Data

The data used for this methodology is the same as that used for the “place” section: all personal datives and dative presentatives that show regional distribution in the South.

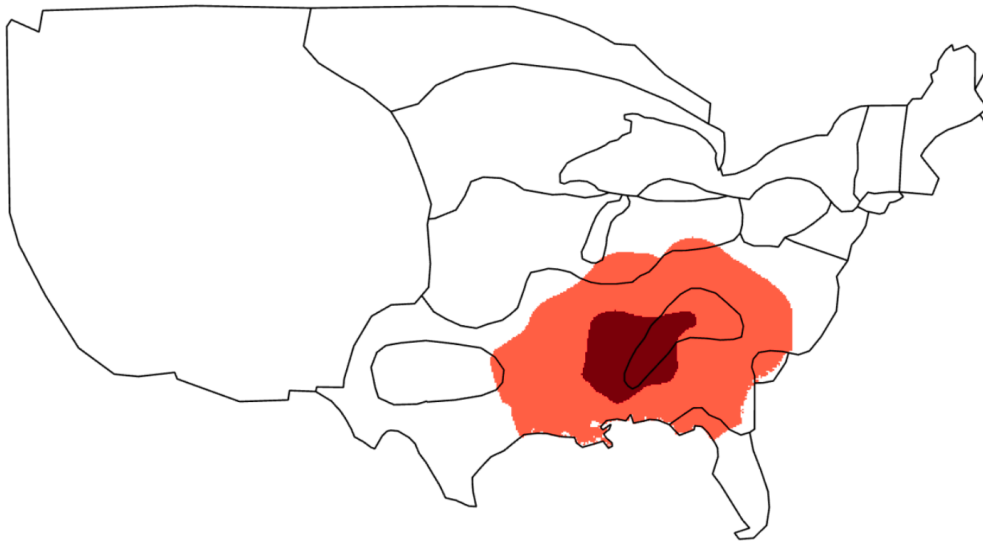


Figure 12. Half of All Sentences Dialect Region w/ ANAE Regions. Lighter Red: 3–4.5, Darker Red: 4.5–5

4.2 General Methodology

The base procedure for creating a dialect region for this aspect of dialect region analysis is based off of the tessellation methodology from “place”. This methodology was chosen because it was the only explored definition of place that divided the country into areas that contained multiple points per area. The procedure was followed almost exactly (Generate Tessellation, Spatial Join, Polygon to Raster, Reclassify, Cell Statistics, Focal Statistics); the only difference between the three definitions of measure is the Overlay statistic for the Spatial Join.

The average definition of measure used the Overlay statistic of Average (so each area would have the average acceptability judgement for all points in that region per sentence). The minimum definition of measure used the Overlay statistic of Minimum (so each area would have the minimum acceptability judgement for all points in that region per sentence). The maximum definition of measure used the Overlay statistic of Maximum (so each area would have the maximum acceptability judgement for all points in that region per sentence).

4.3 Analysis

The resulting maps for the three definitions of measure are shown in Figures 13–15. Their symbologies differ slightly due to the different ways in which the sentence regions layered, but still have about the same proportion of accepted sentences in each specified color category.

When comparing the three maps, we may see that the maximum and average definitions produce similar regions, while the minimum definition produces an inaccurate region, seemingly showing only noise on the map. This noise may be characteristic of the tessellation method and not of the minimum measure definition itself (see Map 8), but it does show up most prominently in the minimum map, so much so that it disregards the true significant regional distribution.

The average and maximum measure definitions both have their strengths and weaknesses. The

Sentence Number	Sentence
F1002	“Here’s you a piece of pizza.”
F1007	“Where’s us some ripe blueberries?”
F1009	“Where are me some country boys?”
F1067	“Where’s me a screwdriver?”
F1068_1	“Okay, now where’s you a pillowcase?”
F1069	“Where’s us a good place to eat around here?”
F1070	“Here’s me a good pair of jeans!”
F1070.1	“Here’s me a good pair of jeans.”
F1071	“Here’s us a gas station – pull over!”
F1098	“Here’s him a nice cup of coffee.”
F1099	“Here’s John a glass of iced tea.”
F1116	“Here’s you some money.”
F1117	“There’s you a piece of pizza.”
F1118.1	“Here are you some books.”
F1119	“Here is you a new bunny.”
F1120	“Here is you some camping ideas.”
F1121	“Here comes you a bus.”
F1122	“Here’s you some fun ideas.”
F1123	“Now there’s me a new Easter dress.”
F1124	“Now there’s us a story.”
F1125	“Now there’s us some easy money.”
F1179	“I’m gonna go and play with me a cat.”
F1180	“I’d put me a marble or two in my pocket.”

Table 4. Dative Sentences

average definition gives a region that has more accurate placement, but that is not very precise; there exist two large core dialect regions in the South where there should be only one. On the other hand, the maximum definition gives a continuous area that makes sense as a dialect region, but that is not accurately placed when compared to the ANAE regions and most of the other maps created in this research.

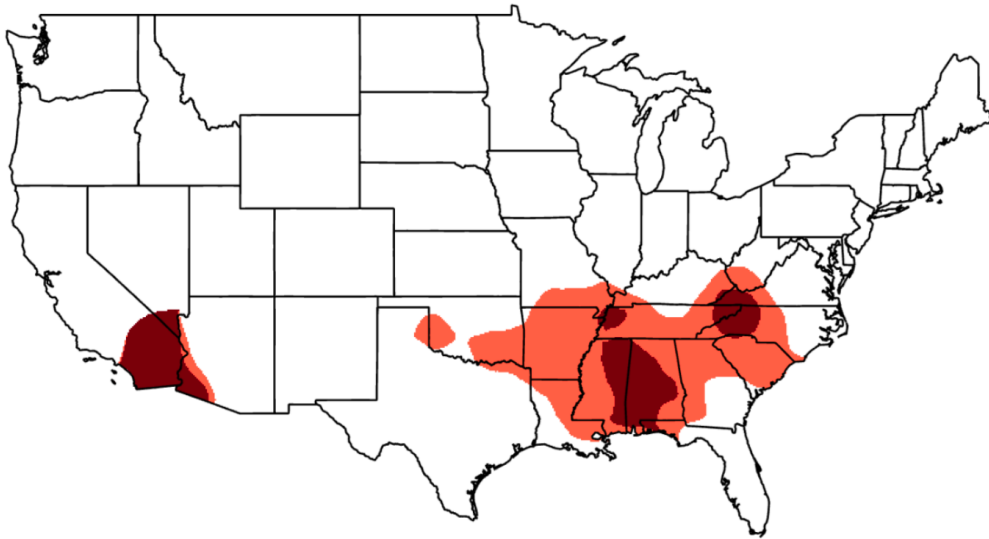


Figure 13. Average Dialect Region. Lighter Red: 9–14, Darker Red: 14–18

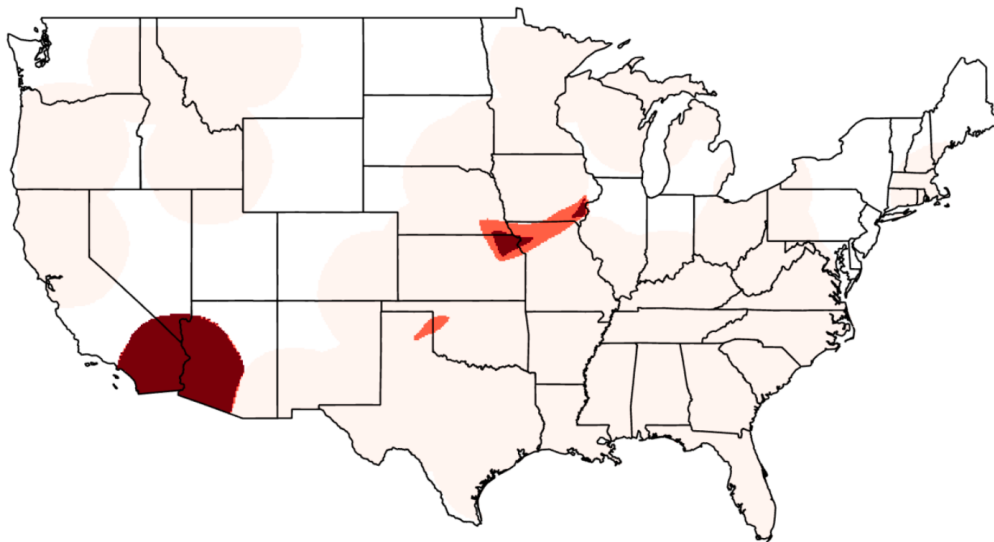


Figure 14. Minimum Dialect Region. Medium Red: 7–9, Darker Red: 9–11

Conceptually, the maximum definition should give a region of the people who dislike the constructions the least, while the average definition should give a region that represents the distribution of acceptability judgements across the country as a whole. Both regions are valid, so there is no clearly superior region conceptually, either. Perhaps we could use these two methods together, taking into account both maps and attempting to draw a dialect region informed by both. If a superior method must be defined for “measure”, it would be the average method. This method

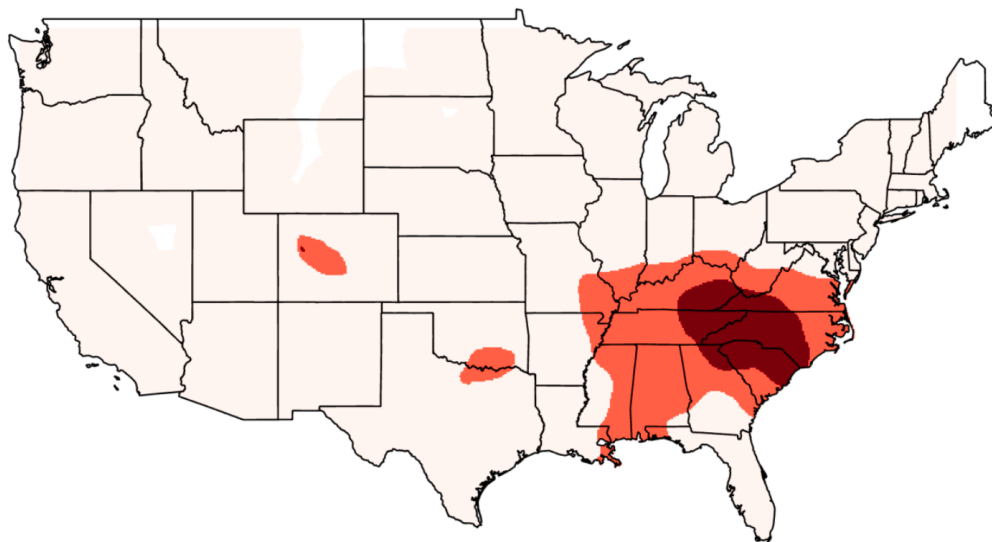


Figure 15. Maximum Dialect Regions. Medium Red: 11–15, Darker Red: 15–18

gives a region more accurate in terms of location, and while the region may not be valid as a dialect region, it still shows more a accurate distribution that can be used to draw an estimated dialect region.

5. Conclusion

In this paper, we have outlined several methods of defining a dialect region from the bottom-up, discussed their conceptual advantages and disadvantages, and compared the results to previously proposed dialect regions. We have suggested that

- (a) the best method of defining a “place” is to use hot spot regions defined on Voronoi polygons
- (b) using at least half of the sentences representing a construction is better than using only two sentences
- (c) the mean judgment of the sentences included is the best way of representing the overall judgment of a construction (as opposed to minimum or maximum)

It is worth noting, however, that these results rely heavily on structure of the YDGP dataset itself, as well as the specific methodologies within the different sections of the paper. While there were certain base methodologies that were used across all methods, other base methodologies could be considered, and there was still variation in methodology within a given section. For example, for “place,” Hot Spots Analysis was used only for the Voronoi polygons, while hand-drawn isoglosses and tessellations used only layering and reclassification. The “best” methodologies mentioned above do not necessarily constitute the definitively superior method of defining a dialect region—there are still many methods to be explored. They are simply the methods that seemed to provide the most reliable results in the present case. So, while these findings should not necessarily be

directly applied to any dialect analysis, they may inform or guide decisions concerning defining a dialect region.

References

- Carver, Craig M. 1987. *American Regional Dialects: A Word Geography*. Ann Arbor: University of Michigan Press.
- Labov, William, Sherry Ash, and Charles Boberg. 2006. *The Atlas of North American English*. Berlin: Mouton.
- Ord, J Keith, and Arthur Getis. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 27 (4): 286–306.
- Sibler, Pius, Robert Weibel, Elvira Glaser, and Gabriela Bart. 2012. Cartographic Visualization in Support of Dialectology. Proceedings Auto Carto 2012, Columbus, OH, 16–18 September 2012.
- Stoeckle, Philipp. 2014. Linguistic Geography and GIS. Available at <https://tinyurl.com/y92lrkd4>. Last accessed June 22, 2020.
- Wood, Jim. 2019. Quantifying geographical variation in acceptability judgments in regional American English dialect syntax. *Linguistics* 57 (6): 1367–1402. doi:doi.org/10.1515/ling-2019-0031.
- Wood, Jim, and Raffaella Zanuttini. 2018. Datives, Data, and Dialect Syntax in American English. *Glossa: a journal of general linguistics* 3 (1): 87.
- Wood, Jim, Laurence R. Horn, Raffaella Zanuttini, and Luke Lindemann. 2015. The Southern Dative Presentative meets Mechanical Turk. *American Speech* 90 (3): 291–320.
- Wood, Jim, Raffaella Zanuttini, Laurence R Horn, and Jason Zentz. 2020a. Dative country: Markedness and variation in Southern dative constructions. *American Speech* 95 (1): 3–45.
- Wood, Jim, Ian Neidel, Sasha Lioutikova, Luke Lindemann, Lydia Lee, and Josephine Holubkov. 2020b. Mapbook of Syntactic Variation in American English: Survey Results, 2015–2019. *Yale Working Papers in Grammatical Diversity* 2 (4).
- Zanuttini, Raffaella, Jim Wood, Jason Zentz, and Laurence R. Horn. 2018. The Yale Grammatical Diversity Project: Morphosyntactic variation in North American English. *Linguistics Vanguard* 4 (1): 1–15.