

Jim Wood*

Quantifying geographical variation in acceptability judgments in regional American English dialect syntax

<https://doi.org/10.1515/ling-2019-0031>

Abstract: The acceptability of many syntactic constructions varies across speakers. In some, but not all cases, such interspeaker variation is due to speakers coming from different dialect regions. However, since neither dialect regions nor acceptability judgments are discrete, it is not always straightforward to determine when variation in a construction's acceptability is geographically constrained. In this paper, I present one method of making this determination using the G_i^* statistic. I apply it to a variety of syntactic constructions, including presentative datives, *so don't I, be done my homework*, intensifiers like *wicked* and *hella*, and *do*-support with *have yet to* sentences. Each shows a different geolinguistic profile, revealing different kinds of issues that quantitative analysis can shed light on. The results show fine-grained geolinguistic differences in syntactic acceptability judgments, demonstrating that the G_i^* statistic can be fruitfully applied to further our understanding of geographically-based variation in acceptability judgments.

Keywords: dialect syntax, geolinguistics, microvariation, American English, regional variation, syntactic variation

1 Introduction

Recent decades have seen an ever-growing interest among syntacticians in what is often referred to as “microvariation,” a term used to describe relatively small differences among otherwise very similar language varieties. Some prominent examples include the research surrounding the SAND (Syntactic Atlas of Dutch Dialects) project on Dutch dialect variation, the ScandiaSyn (Scandinavian Dialect Syntax) and NORMS (Nordic Center of Excellence in Microcomparative Syntax) projects in Scandinavia, and the ASIS project (*Atlante Sintattico dell'Italia Settentrionale* “Syntactic Atlas of Northern Italy”) (Barbiers and Bennis 2007; Poletto and Benincà 2007). There are many more examples (see <http://www.dialectsyntax.org/>), and prominent theoretical work often makes

*Corresponding author: Jim Wood, Department of Linguistics, Yale University, 370 Temple St., New Haven, CT 06520, USA, E-mail: jim.wood@yale.edu

use of lesser-studied dialects in order to illuminate the empirical domain under consideration (see, for example, Zanuttini and Portner 2003; Benincà and Poletto 2004; Harris and Halle 2005; Torrego 2010; D'Alessandro and Scheer 2015). Indeed, Kayne (2005: 283) argues that we should see microvariation as a kind of microscope that allows us to see finer-grained properties of the language faculty than we otherwise might, a view that seems to be shared by a growing number of theoretical syntacticians.

At the same time, there has been an increased interest in using new tools for collecting data. For syntacticians, acceptability judgments form an important part of the empirical base for the study of syntax. While this is often done on a small scale, experimental work that blossomed in the 1990s showed that we can get reliable, replicable results by studying the judgments of non-linguists (Schütze 1996; Cowart 1997; Gordon and Hendrick 1997). Even more recent work has shown that studies conducted entirely online are reliable as well (Sprouse 2011), and more and more studies are making use of crowdsourcing platforms like Amazon Mechanical Turk to administer judgment studies and test hypotheses quickly and efficiently (Ackerman 2015; Erlewine and Kotek 2016; Champollion et al. 2016; Fetters and White 2016).

The goal of this paper is to present some recent results from ongoing pilot work on syntactic variation in North American English. In so doing, we outline a method to investigate geographic variation in acceptability judgments. It can be applied in principle to any set of data that might vary geographically. The method outlined below is designed to answer the question of whether geography matters for the acceptability of a construction, and where a construction is especially likely or unlikely to be accepted.

It does not come with a theory of the task. For example, we cannot guarantee that participants use the sentences that they accept, or that they do not use the ones that they reject. Nor does it provide all the answers to questions one might ask about syntactic variation. Nevertheless, studies of acceptability judgments are already part of the standard set of investigative tools used by linguists. How acceptability judgments relate to the theory of grammar is a broader question, which cannot be answered in any general way in this paper.

The method outlined here does, however, bear on the question of the relationship between acceptability and grammar, and the results can be quite interesting. Knowing that geography matters for the acceptability of some construction has important implications for our understanding of syntactic variation. If the acceptability of a construction varies across speakers, it is possible that it varies due to the speakers having different grammars. But it is also possible that there is another explanation. Processing factors, for example, can introduce a level of variation in acceptability that might not be reducible to grammar (Gibson and

Thomas 1999). If, however, acceptability is constrained by the geographic region of the speakers, a processing explanation becomes much less likely, since it seems unlikely that people from different regions will be subject to distinct processing effects. It is rather more likely, in such a case, that speakers vary because they have different grammars in different geographical regions.¹

But determining whether geography matters for a construction is not always straightforward. For many patterns of dialect variation, it is not a matter of everyone in a region using some form, and everyone outside of the region not using that form. It is often a matter of degree; and geolinguistic borders are generally not rigid (cf. Bart et al. 2013). The same holds for acceptability judgments that vary geographically. The purpose of this paper, then, is to present several examples of sentences whose acceptability varies across speakers, and then outline an efficient way to go from wondering if the acceptability of a syntactic construction is geographically constrained to answering that question. Along the way, we will see several kinds of results that might show up, from cases that are immediately quite clear to cases that are not so clear.

We will discuss the general methodology used in survey construction and administration in Section 2. We will then discuss the “hot spots” analysis that is the primary focus of the paper in Section 3, and illustrate it with two fairly clear case studies. In Section 4, we apply this analysis to three more case studies, each of which brings its own issues and lessons. Section 5 concludes the paper.

2 Methodology

The results reported in this paper come from a series of surveys, administered as part of a larger, ongoing investigation of syntactic variation in American English. Participants were presented with the test sentences and asked to rate each one on a scale from 1 to 5, in which 1 represents a sentence that the participant judged as unacceptable in his/her own informal speech and 5 represents a sentence that sounds perfectly acceptable. The instructions and a sample sentence is provided in Appendix A. The surveys each consist of 45 sentences, including 15 control sentences, 15 filler sentences, and 15 test sentences. The filler and test sentences involve constructions that are known to vary across speakers. The control sentences were used to filter out participants without sufficient English proficiency or understanding of the task. We administered our survey using Amazon Mechanical Turk (AMT), which has been shown to be a viable platform for survey administration

¹ But see Wood et al. (2019) for a discussion of how pragmatics and syntax may interact to affect how widespread geographic variation may be.

(Behrend et al. 2011) and a particularly useful one for research in experimental syntax and semantics (Gibson et al. 2011; Kotek et al. 2011; Sprouse 2011; Karttunen 2014; Erlewine and Kotek 2016). AMT is especially useful for syntactic dialect research because it provides access to thousands of respondents throughout the United States from a wider variety of demographic and educational backgrounds than are generally available to us in university settings (Sprouse 2011). In general, the distribution of participants across demographic categories in our work is consistent with the findings in Ipeirotis's (2010) study of the demographics of AMT users. See Wood et al. (2015) and Zanuttini et al. (2018) for further discussion of the survey design and criteria for inclusion in the final dataset.

3 Hot spots

While some maps of acceptability judgments show very clear geographical patterns, others are more complex. One response to this situation might be to map out only the very clear cases, and take those to be the “real indicators” of geographical syntactic dialect variation. Another response is to find some way to smooth over the data. That is, take many cases, of different phenomena, find cases that pattern together, and generate isogloss bundles (see Grieve et al. 2011 for one systematized way of doing this, which overlaps in part with the method used here).

Both of these are acceptable approaches to the study of dialect geography, and we might learn a lot from them. But they also come with a problem. Any amount of smoothing or bundling may give rise to unrealistic expectations about the discreteness of dialects. It is very common, for example, to encounter even well-trained linguists who say things like, “I was very surprised to find out that a student from PLACE X accepted SENTENCE Y – I thought that SENTENCE Y was only found in PLACE Z.” Many cases of syntactic acceptability are not so clear cut. It is worthwhile to know when geography matters for a construction's acceptability, but at the same time to appreciate that there is variation in many places. In this paper, I describe the use and visualization of a fairly easy-to-use statistical test referred to in ArcGIS as the “hot spots” analysis. The ultimate goal is to capture a picture of geographic variation in the acceptability of a single sentence, combining the raw data, visual statistical smoothing on that data, and statistical significance testing into one map image.

The hot spots analysis has been used in a number of geolinguistic studies in recent years, most prominently in work by Grieve (see Grieve 2009, Grieve 2014, Grieve 2016; Grieve et al. 2011, Grieve et al. 2013), who sees “introduction of local spatial autocorrelation analysis” (including hot spots) as “a major methodological contribution” of his work (Grieve 2016: 140). It has been applied to variation in phonetic/phonological production (Grieve et al. 2013; Grieve 2014), phonological

perception (Kendal and Fridland 2016), morphological form (Tamminga 2013), lexical choices (Grieve 2009, Grieve 2016; Grieve et al. 2011), and syntax (Sibler et al. 2012; Bart et al. 2013; Ko 2016; Stoeckle 2016). The kinds of data used in the studies of syntax, however, are quite different from the present work. Sibler et al. (2012) and Bart et al. (2013) analyze proportions of one variant chosen over another variant (after using Kernel Density Estimation to infer smooth intensity values). Ko (2016) analyzes variation in the attested frequency of different classes of adverbs in different syntactic positions in English corpora from around the world. Stoeckle (2016) studies the results of a translation task, and creates a measurement for each location of how much variation is found in the responses to this task. The hot spots analysis is used to identify areas of high and low variation.

However, the hot spot analysis has not, as far as I know, been applied to the kind of data that syntacticians use most: acceptability judgments. There is no obvious reason why the hot spots analysis could not be applied to acceptability judgments. As pointed out in Tamminga (2013: 117), it is designed to work on “quantitative measures of linguistic variables.” Acceptability judgments, especially when collected with a continuous measurement such as a Likert scale, are certainly quantitative measures that can be (and often are) analyzed with parametric statistics. One of the main points of the present article is to show how variation in acceptability judgments can be measured, analyzed, and visualized geographically. This article should not be taken to suggest that it is not useful to analyze variation in the frequency of a construction (or the preference for one over another). However, as discussed above (and throughout this article), it is useful to know where a given sentence/construction is accepted *and* where it is rejected.

We turn now to the workings of the hot spot analysis. The hot spots analysis identifies statistically significant hot spots and cold spots using the Getis-Ord G_i^* statistic (Ord and Getis 1995). This statistic tells us whether an observed spatial clustering of high or low values is more pronounced than one would expect in a random distribution of those same values. The G_i^* statistic is given in eq. (1).²

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2 \right]}{n-1}}} \quad (1)$$

In this formula, x_j is the value for the feature j – in the present case, the acceptability judgment for a given sentence in a given place. $w_{i,j}$ is the spatial

2 The formulas in this section are taken from the software’s website at <http://goo.gl/K8vsZ4>.

weight between feature i and j . (See below for more discussion of the weight.) n is the total number of features. The terms \bar{X} and S are given in eqs. (2) and (3):

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (3)$$

The resulting statistic replaces the acceptability judgment with what is essentially a z-score, but one that is computed on the basis of the values surrounding the area in question. An area is a “hot spot” if its value and the values of the areas around it lead to a statistically significant z-score.

The spatial weight can be determined in several ways, and there is much discussion in the literature cited above regarding how to select a spatial weighting measure. Kendal and Fridland (2016: 183) note that “there is no foolproof method for choosing the most appropriate spatial weighting function.” In this paper, I follow the default, (software-)recommended method, which is called “Fixed Distance Band” (also used in Grieve 2009; Grieve et al. 2011; Kendall and Fridland 2016; Ko 2016; Stoeckle 2016).³ This can be thought of as a floating window that sits on top of each location, and captures that location and its neighbors within a certain, determined distance, called the critical distance. Neighboring features (i. e. acceptability judgments in particular places) within a specified critical distance receive a weight of one. Features outside of that critical distance receive a weight of zero. Thus, the comparison is one that compares the distribution of values inside the critical distance with values outside of that critical distance. The critical distance can be specified manually, or it can be determined by the optimized hot spot tool on the basis of the best fit for the distribution of the data; essentially, it tries to set a distance that ensures

³ This is sometimes referred to as a binary weighting function. These authors vary somewhat in how they choose the critical distance. In contrast, Tamminga (2013), Grieve et al. (2013), and Grieve (2014, 2016) use a reciprocal weighting function, where instead of receiving weights of 1 or 0, a range of weights are assigned so that closer points get progressively higher weights than points further away. For this measure, it is important to set the power level, and again, there is no foolproof way to choose. See Grieve (2009: 157–158) for points in favor of fixed distance “binary” weighting function, and Grieve (2016: 114–118) for points in favor of a reciprocal weighting function.

that each feature will have at least one neighbor (ideally a few), and that no feature has too many neighbors (e. g. the whole data set). The optimized hot spot tool was chosen in the present case, since there was no particular reason to select a critical distance *a priori*.⁴

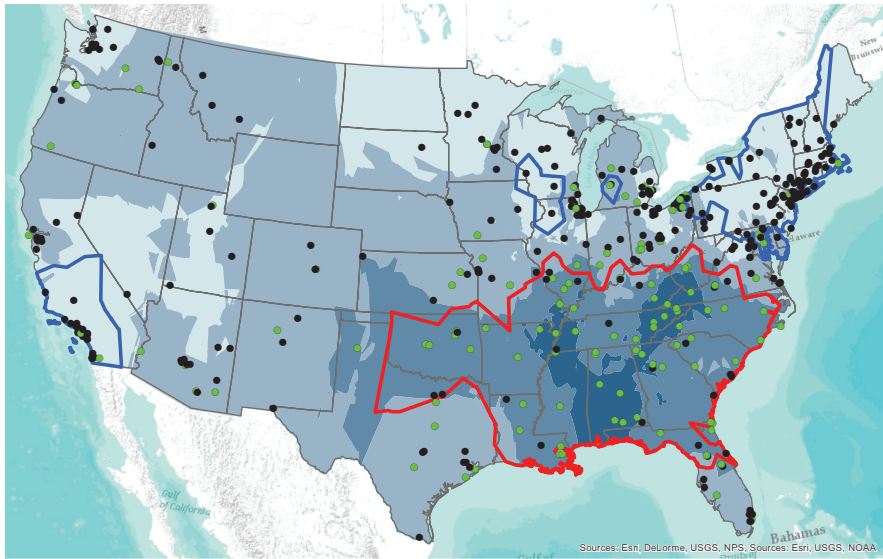
It should be clear that a variety of choices can be made in using the hot spots analysis tool. I have indicated here which choices I have used in the case studies below, for nationwide acceptability judgments. (Fixed distance band, with automatic determination of the critical distance; see Appendix B for specific information about each map.) It is possible that other choices would be useful for different kinds of questions. However, it is my hope below to show that this choice yields insightful, sensible results for the kind of data under consideration. I will illustrate this by showing several cases that pick out exactly what we would expect based on a cursory view of the data and prior expectations of what regions are expected to be relevant for those cases. I will then show how even when the data looks quite variable, visually, the hot spots analysis is able to pick out sensible results.

First, consider the dative presentative construction (Wood et al. 2015, Wood et al. 2019). The map in Map 1 shows the distribution of acceptability judgments for the sentence *Here's you some money*. In this map, judgments of 1 and 2 are grouped together and represented as black dots. Judgments of 4 and 5 are grouped together and represented as green dots. Judgments of 3 are left off of the map entirely.⁵ I hasten to add that this is for visual presentation only – all of the statistical analyses discussed in this paper include all of the judgment data (including 3s), and do not in any way “collapse” distinct values (so 1s are distinct from 2s).

Map 1 clearly reveals a distribution wherein people from the South accept the sentence and people outside of the South by and large reject it. (Note that the points on the map indicate not where this person currently lives, but rather where the person was primarily raised as a child; see Zanuttini et al. 2018 for detailed discussion.) The hot spots analysis picks out exactly what we would expect it to pick out: a series of contiguous hot spots in the South and a series of contiguous cold spots in the Northeast (and California). I will refer to contiguous sets of hot and cold spots as hot/cold spot regions. The absence of hot spots in the other areas indicates an average level of variance in those

⁴ An equally sensible strategy is to select one distance and apply it across all examples for consistency. Thus, Grieve (2009) chooses a 500 mile critical distance for all variables in his data, while Grieve et al. (2011) select different distances for different variables.

⁵ This is one reason that the number and distribution of data points varies across maps in this paper. Since 3s are not shown, a data point may be present but not visible in one map (since it is 3 for that sentence) and yet visible in another (since it is not a 3). The other reason that number and distribution of data points varies is that different maps may come from different surveys.



Map 1: Here's you some money.

areas. The other factor to consider is transition zones. The area between the hot spot region and the cold spot region contains neither hot spots nor cold spots. This is because of how the hot spots analysis works. Recall that it essentially floats a window over each place and determines if the value of that place and the other places within that window are significantly high or low. Transition zones, then, are generally not going to be hot or cold, because the window will cover areas with high acceptance and areas with low acceptance. This translates into high variance with a moderate mean – not especially hot or cold. For phenomena that have a sharp border between them, that border will generally not show up as a border between immediately adjacent hot and cold spots. In fact, the hot spot borders in Map 1 are fairly close to that, considering how the hot spots analysis works.

With all this in mind, the borders drawn around hot/cold spot regions, in many cases, should not be considered isoglosses of dialect regions, even if they sometimes come close to this. They only tell us where the relevant hot spots are – where judgments are especially high or low. The values themselves are still important in determining the linguistic nature of the phenomenon and its distribution.

Two further remarks on the maps are in order. First, the shaded background reflects a visualization technique called interpolation. What interpolation does is

estimate, for any given spot on the map, what its value is likely to be, on the basis of some specified algorithm. In the present case, it takes the 12 nearest neighboring values and applies an inverse-distance weighted algorithm, so that closer values have a greater effect on the result than the values farther away. This technique is not a test for statistical significance but is used here purely for visualization purposes. It can be very helpful in interpreting the spread of values in a dataset.⁶

Second, the borders around the hot/cold spot regions are drawn using Voronoi polygons. A Voronoi polygon is a polygon drawn around a point. Within that polygon, every space is closer to that point than to any other point on the map. In this way, the geographical space is partitioned completely around all of the points. These polygons serve as the basis for the borders around the significant hot/cold spots.

The distribution in Map 1 is useful in showing that the hot spots analysis can capture rather large geographic areas, when that is truly the level of granularity involved in a particular geolinguistic distinction. That is, in Map 1, it picks out the entire southeast, and visual inspection of the judgment values on the map suggests that this is exactly correct. However, it is worth noting that the formula does not need to pick out such a large region. The large region is actually just the amalgam of a series of individual locations, each of which is considered a hot spot (or cold spot, for the Northeast or California). So it is worth pointing out that even with the entire U.S. as an input, the hot spots analysis can pick out rather small areas, if that is what the data shows.

To show this, consider the *so don't I* construction, which is exemplified by sentences such as (1).

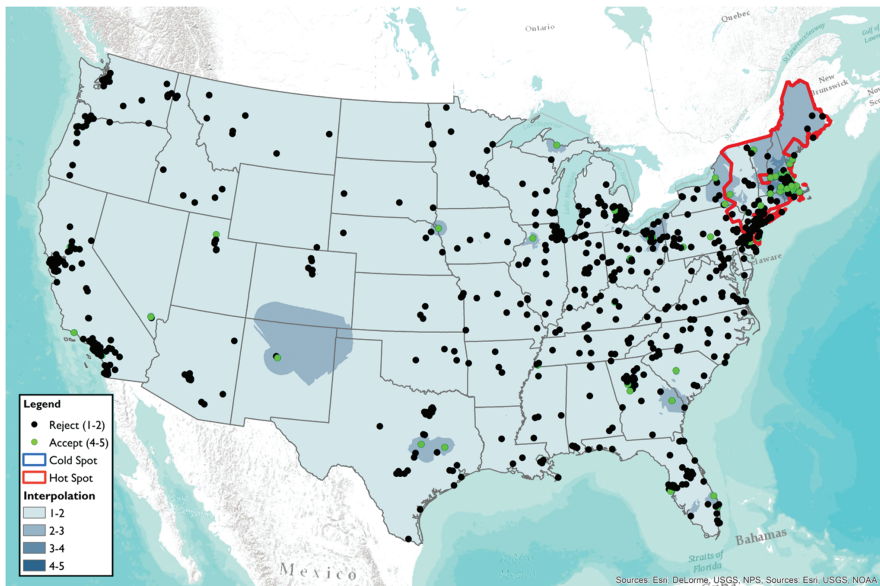
- (1) *Sure I could help you, but so couldn't my brother, and he's free right now.*

In this sentence, the phrase *so couldn't my brother* is actually truth-conditionally equivalent to *so could my brother* (but see Wood 2014 for discussion of some pragmatic nuances found in the construction). Previous research has indicated that it is characteristic of eastern New England, especially Boston, Massachusetts (see Wood 2014 for references). However, this has never been confirmed in a nationwide study.

Map 2 presents the results of the sentence in (1). It shows that the hot spots analysis picks out exactly the basic region we would expect, the Northeast. Visual inspection of the judgment values represented in Map 2 confirms that this is basically the correct result. While one does find the occasional one-off person outside of New England accepting the construction, it is only in New England

⁶ But see Wood et al. (2019), where interpolation is put to analytical uses that go beyond just visualization.

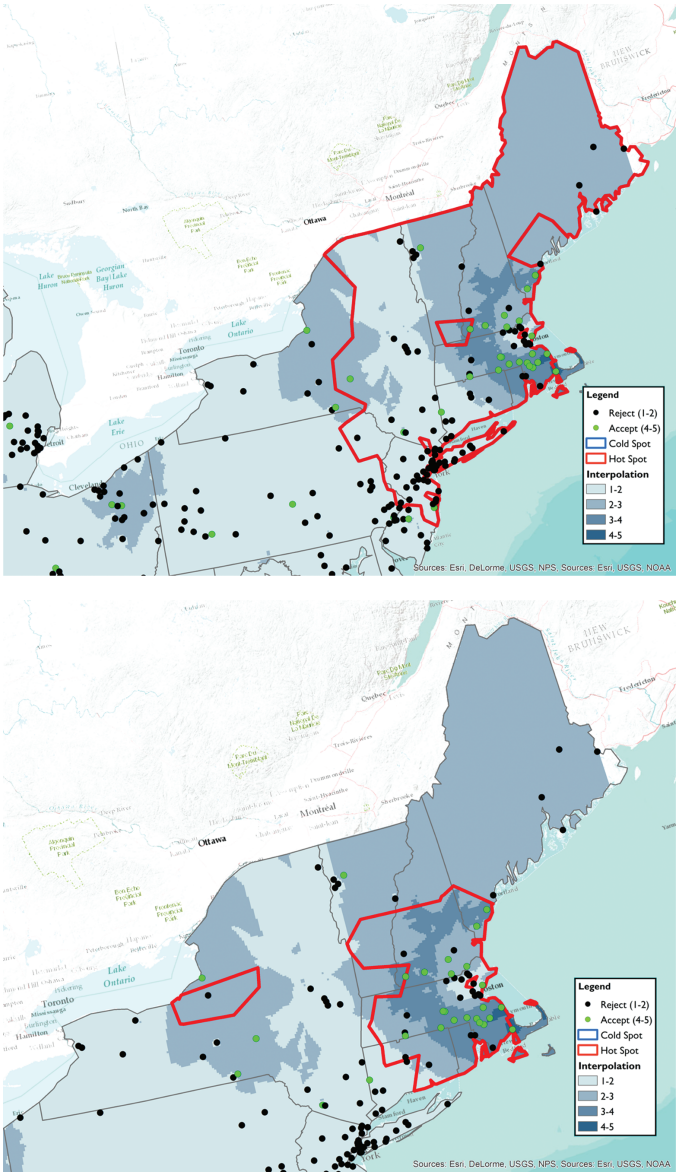
that one finds a substantial number of people accepting the construction. However, the hot spot region is also rather large, and extends further west than one might expect on the basis of the values of the judgments themselves.



Map 2: Sure I could help you, but so couldn't my brother, and he's free right now.

This is because the critical distance determined by the optimized hot spot tool for this data set is rather high – approximately 501 km. The western area is falling under the influence of eastern New England, where there are many acceptances within the 501 km window. Given the generally low judgment of the sentence across the dataset (mean = 1.53, SD = 1.02), this is enough to yield a high z-score on those points. When the critical distance is set manually to 150 km, the area picked out is correspondingly much smaller, and basically covers the darkest areas of the interpolation shading.⁷ Map 3 shows a comparison between 501 km and 150 km in the Northeast region.

⁷ As noted earlier, there is no getting around the fact that a critical distance has to be chosen, and in some cases, it makes sense to use different settings for different examples. Thus, Grieve et al. (2011: 12) say that “to identify spatial clustering in the values of individual variables as accurately as possible, it is important to fit the spatial weighting function for each variable.” Their critical distances range from 200 miles to 1000 miles. Referring to the critical distance as the “cutoff distance,” they continue: “The cutoff distance essentially sets the level of resolution for the analysis. A smaller cutoff is better for identifying smaller clusters, whereas a larger cutoff is better for identifying larger clusters. Setting the cutoff distance is problematic,



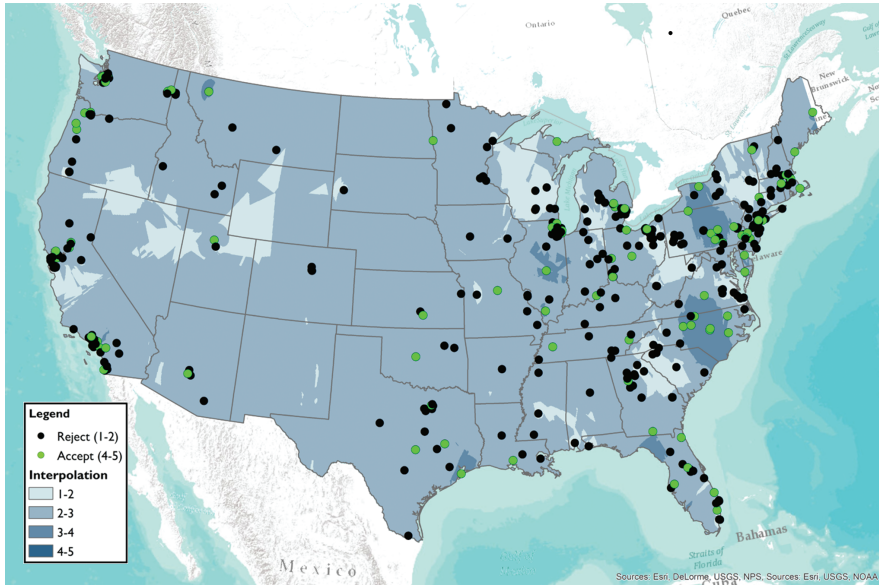
Map 3: “... so couldn’t my brother ...” 501 km (Top) vs. 150 km (Bottom) critical distance.

however, because it is possible for different linguistic variables to exhibit regional patterns at different levels.” See also Ko (2016: 22) for further discussion of this point.

Notice also that there are no cold spots in Map 2. This is because outside of New England, almost all areas are uniform in giving very low ratings to the sentence. There is no area where the rating is especially low – all areas are low. If there were a substantial number of areas that had some amount of variation and others that did not, then there might be cold spots. This point is worth considering in the interpretation of both hot and cold spots – they are relative to the entire data set. To take a hypothetical example, consider southern California, where judgments of *so don't I* are quite low. (Note the cluster of black dots in Map 2, as well as the light blue color of the surrounding area.) It is not a cold spot because it is no lower than most of the rest of the country. Now suppose that outside of southern California, we changed the judgments so that they all varied between 2 and 4. Without changing the values in southern California itself, it would become a cold spot. The point is that from one kind of linguistic perspective, our view of California should not change – people there reject *so don't I*; the construction is not characteristic of that region. It is only in comparison to the rest of the country that our view may change – with respect to the question of whether southern California is any different from the rest of the country, or the broader question of whether geography makes a difference in the acceptability of the construction.

Before concluding this section, I wish to note that the hot spots analysis does not always find hot or cold spots. (It would be worrying if it did, of course, because the cases where we find results would then be less interesting.) Quite the contrary, a number of cases we have studied yield no results at all. Consider, for example, Map 4. This map shows the acceptability of subject control past an object (underlined), as in *John_i threatened me PRO_i to come to my house*, a sentence recently noted to vary across speakers by Hartman (2011: 127). The map in Map 4 indicates that this variation is found everywhere. That is, in nearly every location (at the appropriate level of granularity), there are some speakers who accept the sentence, and some who do not. There is no evidence that it is conditioned geographically. The hot spots analysis is consistent with this, picking out no area as statistically significant.

Map 4 also shows why it is useful throughout for maps to show both the interpolation and the hot/cold spot regions. Interpolation provides a clear visualization of the data, but with no indication of statistical significance: there is always a result. Even in Map 4, where there is no reason to think that the variation is geographically conditioned, we see the interpolation throughout the country. Hot spot testing, on the other hand, gives an indication of statistical significance, but in many cases does not provide a very good visualization of the data, and provides no visualization for patterns that do not reach statistical significance. On the other hand, interpolation also provides information about the nature of the judgments: where they are in the 4–5 range, where they are in the 3–4 range, etc. The hot spots test does not do this: it only tells you where the values are higher than one would



Map 4: John threatened me to come to my house.

expect by chance – the values themselves could be on the low side (even a 2–3 average could be a hot spot, if the rest of the data were quite low), or on the high side. For this reason, we think it is useful to include both kinds of analysis: interpolation for visualizing the pattern, and hot spots testing for statistical significance. They may present slightly different pictures at times, but they cannot contradict each other, because they are providing different kinds of information.

In sum, this section has outlined the hot spots analysis, and given some examples of how it works. So far, however, the examples have been relatively clear. The areas picked out by the hot spots analysis are essentially what one would have guessed just by looking at the raw distribution of the judgments. In the next section, we will review cases that are less clear at first glance, and show how the hot spots analysis allows us to appreciate the geographical component of variation in acceptability judgments, without smoothing over the data in a way that obscures that variation.

4 Case studies

As noted above, the case for claiming a geographical component for presentative datives and the *so don't I* construction is relatively straightforward, both visually

and statistically. Visually, it is immediately clear that geography matters for the acceptability of the constructions; statistically, the hot spots analysis picks out basically exactly the region one would expect from a cursory visual inspection. In other cases, we have found, the results are much less straightforward. It is for these cases that the hot spots analysis is most useful, since it can tell us something that a visual inspection cannot. In this section, we review three such cases.

4.1 The *be done my homework* construction

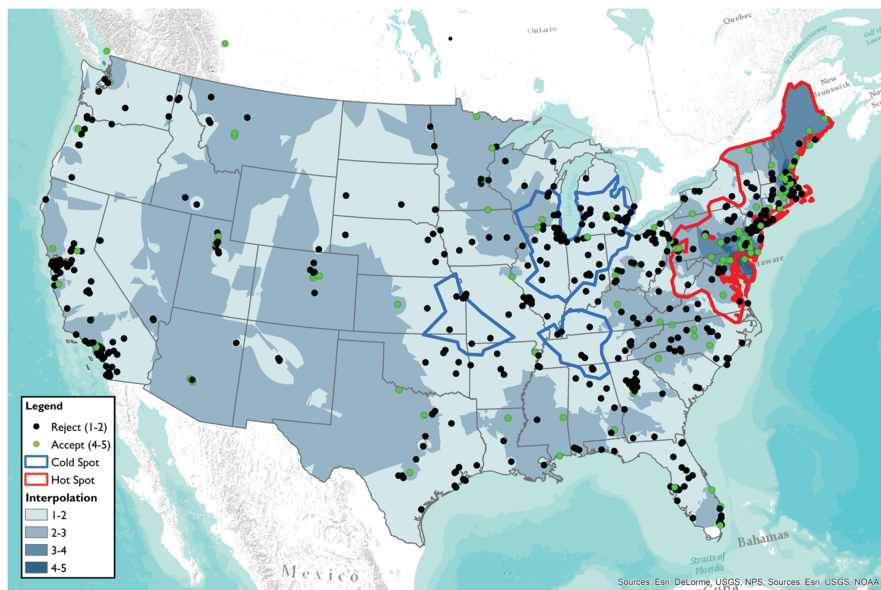
The *be done my homework* (BDMH) construction is exemplified by the following sentences:

- (2) *I'm done my homework.*
- (3) *Are you done/finished your homework?*
- (4) *When I'm done dinner, I like to watch TV.*

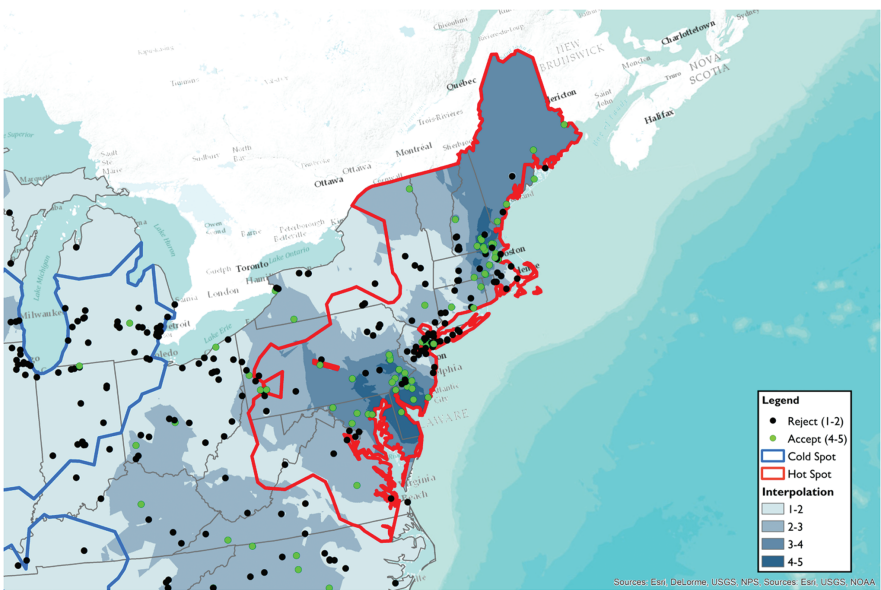
At first glance, it appears to involve omission of the preposition *with*. In fact, inserting *with* in a BDMH sentence after the participle (e. g. *done/finished*) generally yields a grammatical sentence with the intended interpretation. However, it should be noted that the same does not apply the other way round: as emphasized by Fruehwald and Myler (2015), there are many *done/finished with X* sentences that cannot correspond to a BDMH sentence. Running through the set of available analyses, they argue that for many speakers at least, the only viable analysis takes *done/finished* to be an adjectival participle that takes a DP complement directly.

As for its geographical distribution, the BDMH construction has been claimed to be a general feature of Canadian English, but has also been reported in Philadelphia and Northern Vermont (Yerastov 2010). Yerastov (2010: 76) also finds some indication of its existence in Massachusetts, which he attributes to its close proximity to Canada and Vermont. Other than that, he reports it to be “marginal” in U.S. English.

Several BDMH sentences were included in the surveys. We focus on the sentence in (2), which is shown in Map 5. The first thing to notice about the map is that indeed, as Yerastov reported, the construction seems to be marginal in the U.S.: the vast majority of participants rejected the sentence. However, it is not immediately clear whether the acceptability judgments are geographically constrained. People who accept the construction seem to be sprinkled throughout the map. The hot spots analysis, however, picks out the northeast corridor as a hot spot region.

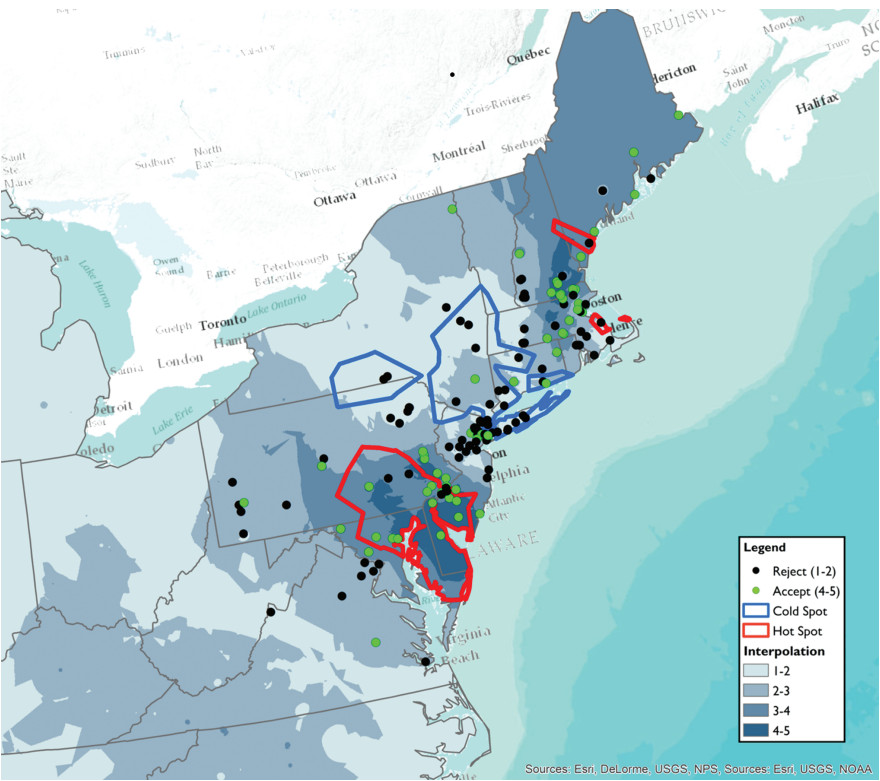


Map 5: I'm done my homework.



Map 6: I'm done my homework. (Northeast closeup).

Zooming into that region, we can see more clearly in Map 6 that there are a lot of acceptances in that area, and considerably more rejections even in the area immediately outside that region. However, it is also easy to see that the region is not homogeneous: there are also many rejections. In this case, we can take the participants within the hot spots region as a separate population, and run the hot spots analysis again – this time including only the people who fall within the original hot spots region. When we do this, we find that indeed, there are other, smaller hot spot regions within the broader hot spot region, and one cold spot region as well. This is shown in Map 7.



Map 7: *I'm done my homework.* (Northeast closeup, hot spot region only).

The hot spot region covers Philadelphia, exactly as one might have expected, given previous descriptions. However, it also covers Delaware, eastern Maryland, southern New Jersey, and a large portion of southeastern Pennsylvania. Visual inspection confirms that this is not a matter of the hot spots analysis being too

coarse grained: the region covered is dark shaded and contains a clear cluster of acceptances, and very few rejections. The cold spot covers southern New York (including Long Island) and much of Connecticut. We see very clearly the contrast between the cold spot region, where there are many rejections and not many acceptances, and the hot spot regions, where it is the other way round.

The smaller hot spot regions found in New England are possibly telling as well. We see that around them, there are many acceptances, especially in southern New Hampshire and eastern Massachusetts, and the interpolation behind it gives a very dark shade, indicating high acceptance in the area. In this case, the fact that more of the region is not statistically significant may have to do with granularity of the parameters of the test. When the critical distance is set manually to 100 km, most of the dark shaded area in eastern New England is picked out as a statistically significant hot spot (not shown). This is also presumably why the smaller hot spot regions have only black dots; with a large critical distance, the more important fact about each spot is the judgments surrounding it, rather than the judgments of the spot itself.

The results thus confirm that *be done my homework* is found in the Philadelphia area and in New England, but also show it to be more widespread in those areas than we might have expected. However, it is important to keep in mind that the results are gradient; the construction can be found outside of these regions. We also find layering – hot spots within hot spots, and even cold spots within hot spots. From a methodological standpoint, this is worth keeping in mind, because a hot spots analysis is conducted on the basis of the entire sample. If it picks out a heterogeneous region, it may be instructive to run the analysis again, restricting it to that region, to get finer-grained results. The analysis may not draw perfect isoglosses, but it provides a very useful way of navigating the linguistic terrain.

4.2 Intensifiers *wicked* and *hella*

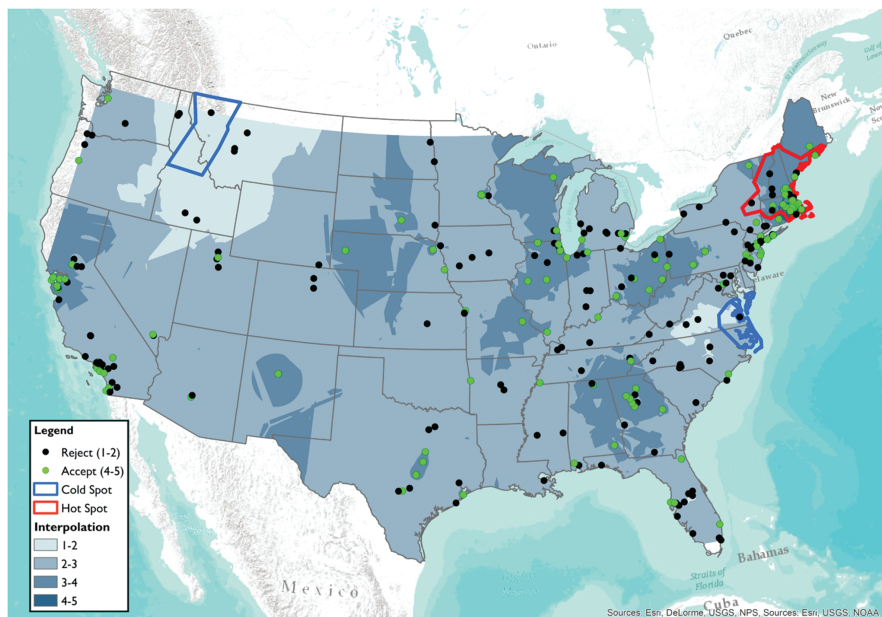
Intensifiers have long attracted the attention of linguists of all sorts. Recently, they have come under much closer scrutiny by syntacticians (Kayne 2005, Kayne 2010; Irwin 2014; Boboc 2016). They turn out to be of substantial interest, due especially to their somewhat idiosyncratic syntactic distributions and selectional constraints. In this section, we will look at how participants responded to sentences with the intensifiers *wicked* and *hella*, which have been long thought to vary across American English dialects. The word *wicked* is used as an intensifier in sentences like the following:

- (5) *Jessie likes that band a wicked lot.*
- (6) *Jamie said that he's been wicked tired lately.*
- (7) *Jordan wants to go there wicked bad.*

It is often interchangeable with mainstream English intensifiers *really* and/or *very* (as in (6) and (7)), but this is not always the case (as in (5)).

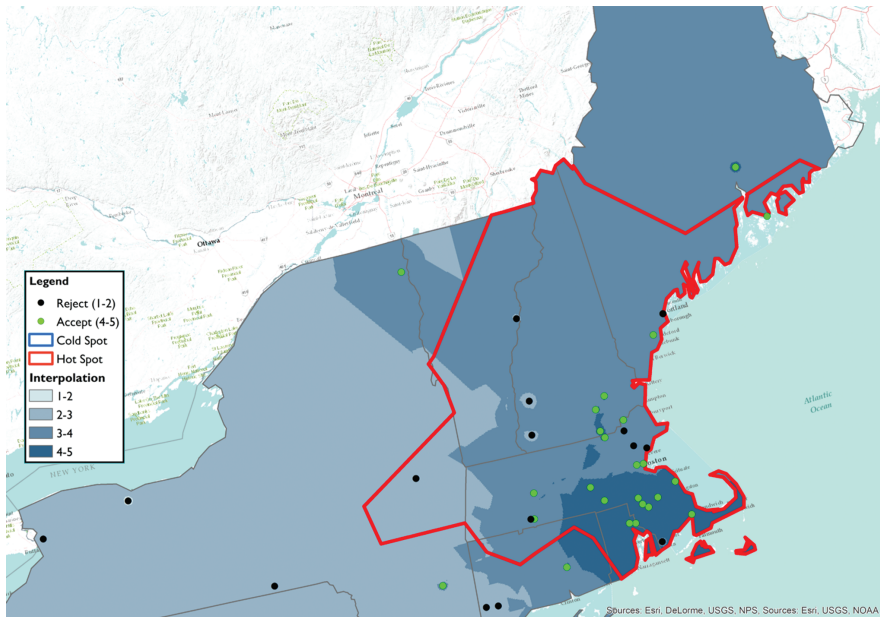
Intensifier *wicked* has long been associated with New England, especially Boston, and is a marker of regional identity. Thus, it appears as such in movies like “Good Will Hunting” as a way of distinctly locating the events of the movie in Boston, and appears as the slogan for the store Newbury Comics (“a wicked good time”), which even in its title celebrates its origins on Newbury Street in Boston. There have been few actual linguistic studies of intensifier *wicked*, but what has been done would seem to support this view. The *Dictionary of American Regional English* (Cassidy and Hall 2013), for example, lists intensifier *wicked* as “New England, chiefly Maine, Massachusetts.” Ravindranath (2011) studied 351 overheard examples in southern New Hampshire.

We start first by looking at the survey results for sentence (5). This sentence is of particular interest because of its *wicked*-specific syntax, which cannot be



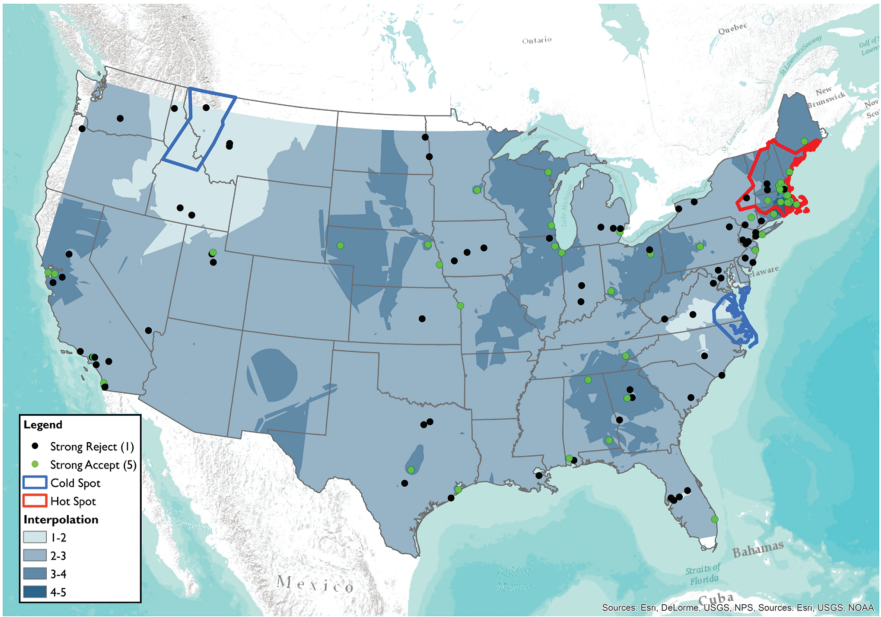
Map 8: *Jessie likes that band a wicked lot.*

readily processed simply by substituting *really* or *very*. Consider the map in Map 8. Here, we see that although it is true that many people seem to accept the sentence in New England, we in fact find acceptances all over the map. There are even a number of areas that seem to contain more green dots than black ones. When the hot spots analysis is applied, however, it picks out exactly the region we would have expected in advance: New England.

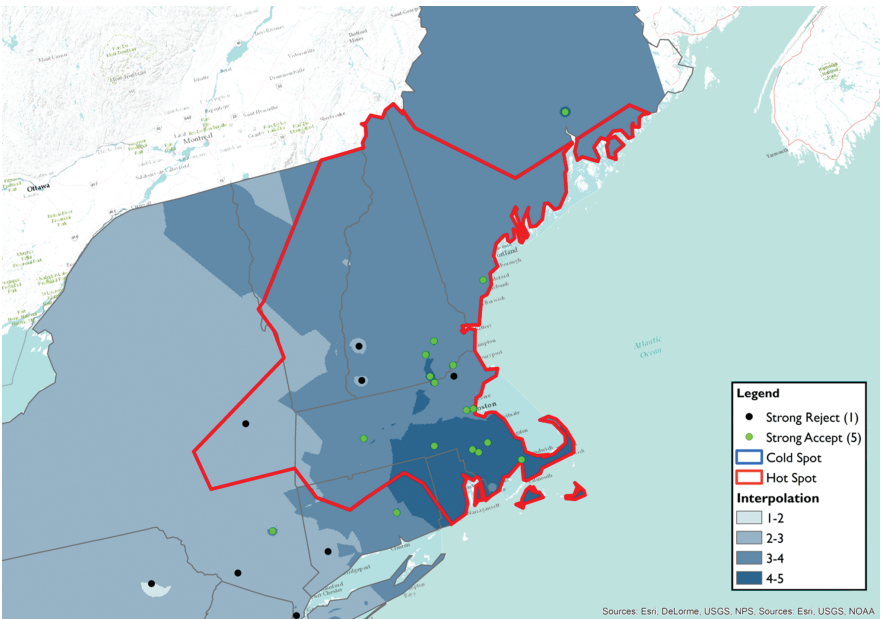


Map 9: *Jessie likes that band a wicked lot.* (Hot spot closeup).

When we look closer at that region (see Map 9), we find that indeed, there are far more acceptances there than rejections. The dark blue shading of the interpolation also indicates a cluster of high values. However, it is nevertheless not obvious through visual scanning of the entire map in Map 8 why that is the only region picked out. It is here where it becomes important that we are collapsing ratings of 4 and 5 and ratings of 1 and 2. In Map 10, we retain only the 1s and the 5s. There, we see that in almost every region of the country, there are no clusters of green dots to the exclusion of black dots. The only exception is in New England (see Map 11). It is in New England where we find an overwhelming majority of people who judge the sentence as a 5 – the highest possible value, and very few people who judge the sentence as a 1 – the lowest possible value.

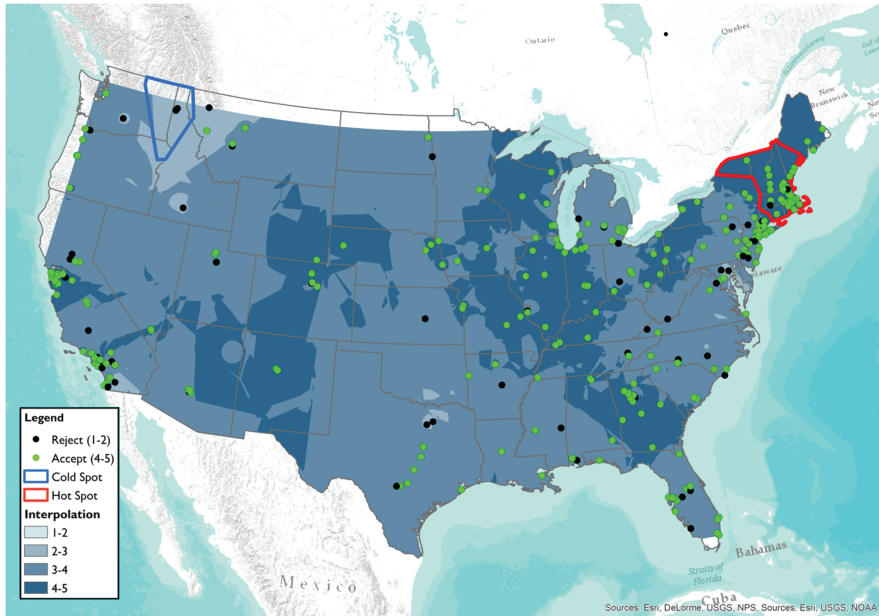


Map 10: *Jessie likes that band a wicked lot. (1s and 5s only).*



Map 11: *Jessie likes that band a wicked lot. (Hot spot closeup) (1s and 5s only).*

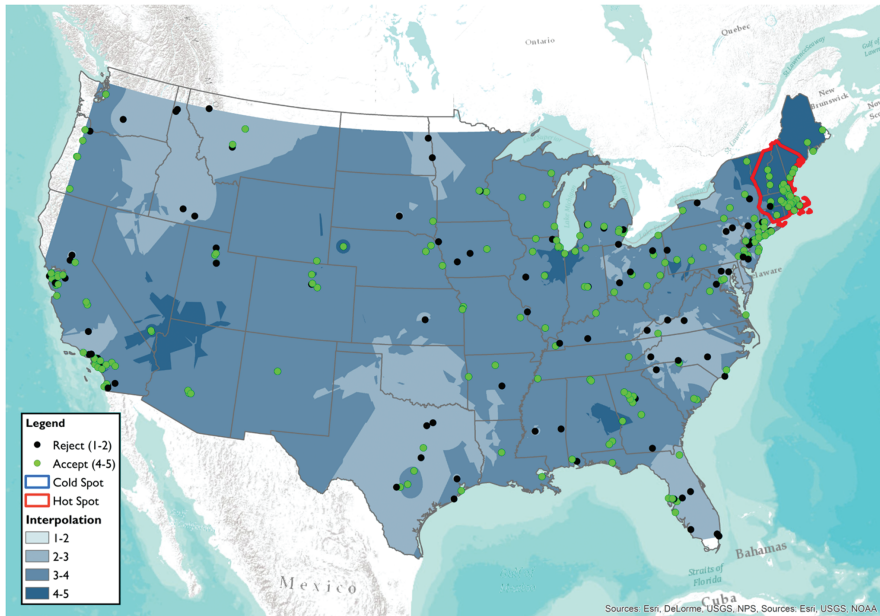
As shown in Maps 12 and 13, this result is replicated with the other two sentences.



Map 12: Jamie said that he's been wicked tired lately.

In sum, with *wicked*, more people across the country seem willing to give “intermediate” positive or negative judgments (i. e. 2 or 4). Mapping it the way that we previously did, collapsing 1/2 and 4/5, makes the expression look more widespread than we might have expected. The hot spots analysis tells us where to look more closely, and finds *exactly* the region we would have expected: New England. It is in New England where people overwhelmingly judge *wicked* sentences as fully acceptable. This illustrates the usefulness of hot spots as a tool for exploring gradient judgment data, where the patterns might not be readily visually observable, but are nevertheless present.⁸

⁸ A reviewer suggests that this is because with *wicked*, we are dealing with lexical variation rather than syntactic variation. While this might be part of the picture, there are at least two reasons to be skeptical. First, the *wicked* sentences under discussion are not just lexically distinct, they are syntactically distinct as well: in a *wicked* lot, *wicked* cannot be replaced by other intensifiers (**a so lot*, **a very lot*, **a really lot*). Second, I have in my research come across other marginal cases that are more obviously syntactic. One example is the data below on *do*-support with the *have yet to* construction (see also Tyler and Wood 2019). However, the present point is unaffected by whether



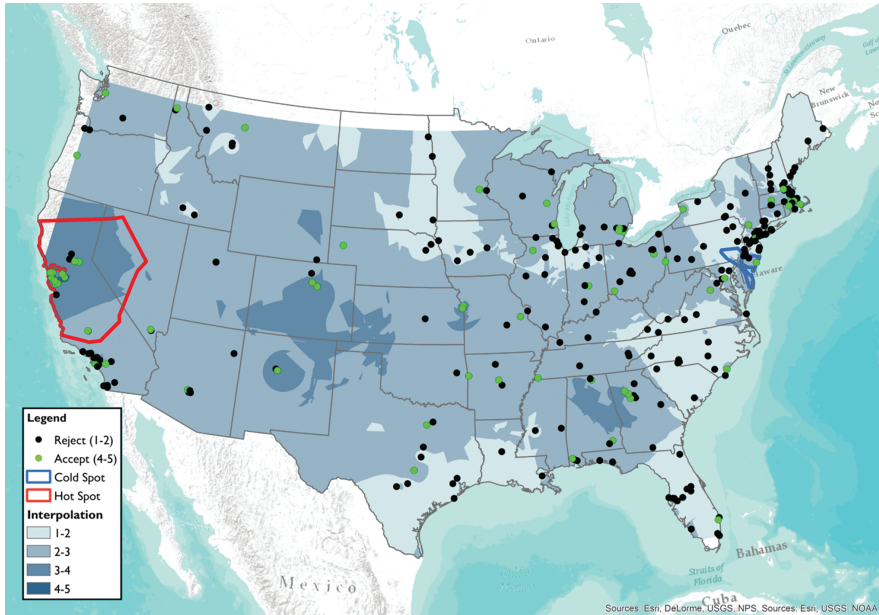
Map 13: Jordan wants to go there wicked bad.

We find a similar result with respect to *hella*. Like *wicked*, it is often interchangeable with *really/very* (as in the example in (8)), but this is not always the case (as in the examples in (9) and (10), which were originally brought to our attention by Wellesley Boboc and Stephanie Harves; see Boboc 2016).

- (8) *That girl is hella smart.*
- (9) *I spoke Spanish for the first time in hella days.*
- (10) *This seat reclines hella.*

Intensifier *hella* is generally perceived to be a California phenomenon (Bucholtz 2006; Bucholtz et al. 2007; Russ 2013). It originated in northern California and is socially salient; within CA, it is seen as a “shibboleth separating the two major regions of the state” (Bucholtz et al. 2007: 343).

we call some variation syntactic or lexical: our primary goal is to illustrate how the hot spots analysis can help determine if a phenomenon is affected by geography.

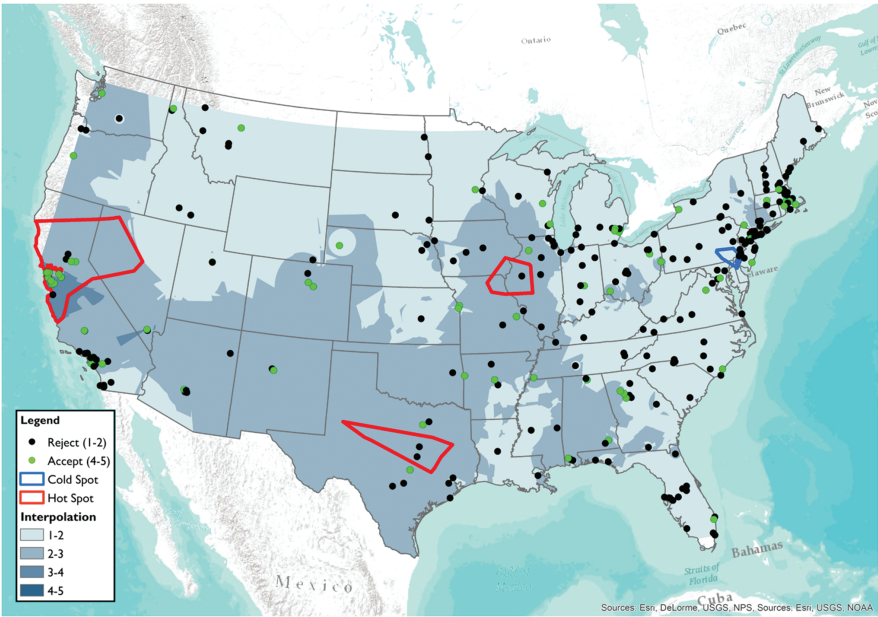


Map 14: I spoke Spanish today for the first time in hella days.

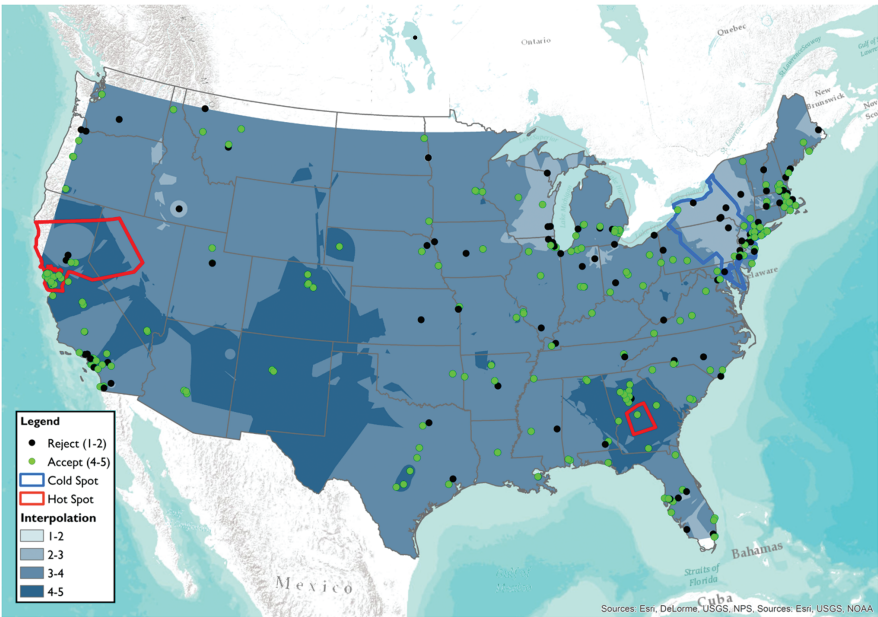
We start by looking at the results for the sentence in (9). (As before, we choose this sentence because of its *hella*-specific syntax.) Consider Map 14. As before, we see a cluster of acceptances where we expect them, in northern California, but we also find acceptances spread across the country. And once again, when the hot spots analysis is applied it picks out exactly the region we expect. We find a cluster of acceptances in the San Francisco Bay area, and not many rejections. As before, when we include only the 1s and the 5s on the map, we find that this is the only area with a cluster of 5s, to the exclusion of all else. As shown in Maps 15 and 16, this result is replicated for the other two sentences: both are accepted across the country, but both isolate northern California as a hot spot region.

Note that both Maps 15 and 16 have one to two other isolated hot spots as well, but these areas do not show up consistently, and whether they are false positives or tell us something new about those regions will have to be left for future research.

To sum up the discussion of intensifiers, in this case, we find that the hot spot analysis is able to detect a pattern through gradient data, and replicate our previous understanding of dialect variation. Not only this, it also tells us something about how people are approaching the judgment task. People are apparently willing to give positive judgments to intensifiers that are not strongly



Map 15: This seat reclines hella!.



Map 16: That girl is hella smart.

associated with their own dialect. (See Zanuttini et al. 2018 for some discussion of possible variation in how participants interpret the acceptability judgment task.) In this case, judgments of “fully acceptable” appear to tell us more about the geographic distribution than in other cases.

4.3 The *have yet to* construction

Most (if not all) dialects of English have some version of the *have yet to* (HYT) construction, which is illustrated in (11) below.

- (11) *John has yet to impress his classmates.*

This sentence means ‘John hasn’t impressed his classmates yet’ (despite the lack of an overt negation marker). Despite the (near) ubiquity of the construction across dialects, however, when one delves into the internal syntax of the HYT construction, one finds a great deal of variation (Bybel and Johnson 2014; Harves and Myler 2014; Tyler and Wood 2019). In this section, we will focus on the co-occurrence of *do*-support with HYT.

The relevance of *do*-support has to do with the status of *have* in HYT: is it a main verb, or an auxiliary? English has both a main verb *have* and an auxiliary *have*. Main verb *have* allows *do*-support, as illustrated in (12), while auxiliary *have* does not, as illustrated in (13).

- (12) a. *John has a car.*
b. *Does John have a car?*
- (13) a. *John has been there.*
b. **Does John have been there?*

The question that arises is which kind of *have* occurs in the HYT construction: the auxiliary, or the main verb?

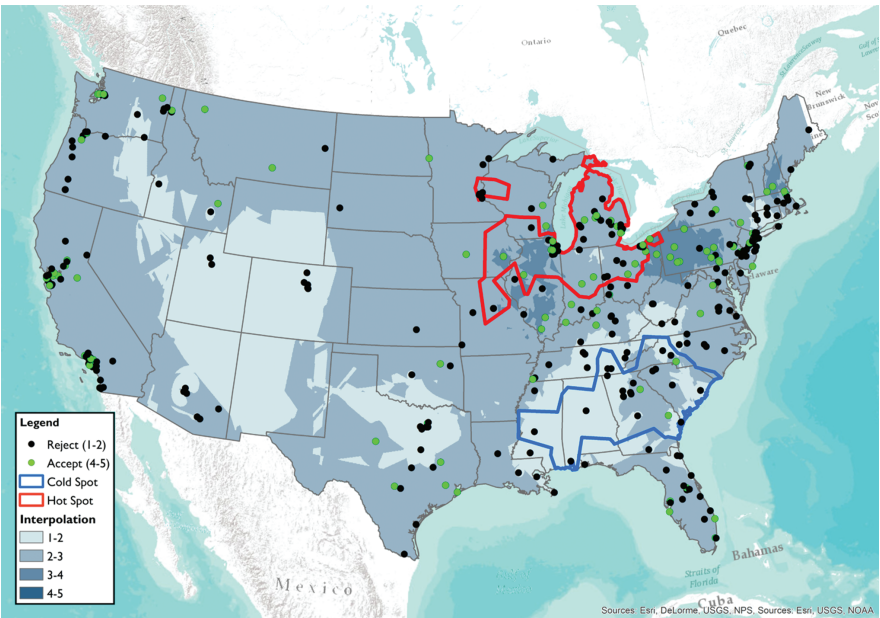
It turns out that there is some variation across speakers. Tyler and Wood (2019) report that the majority of speakers treat *have* as an auxiliary, while a smaller number of speakers treat it as a main verb. Interestingly, however, these are not two different groups of speakers: speakers who accept *do*-support with HYT (showing that they may treat *have* as a main verb) overwhelmingly also allow (or even prefer) *have* to be treated as an auxiliary. This asymmetry goes only one way; there are many speakers who accept *have* as an auxiliary but reject all instances of *do*-support. Tyler and Wood (2019) argue that this

asymmetry should follow from the structural analysis of the HYT construction. For now, we set this issue aside.

What is of interest presently is the geographic distribution of the speakers who accept *do*-support. We included the following four sentences in our surveys:

- (14) *Does John have yet to win the hearts of his classmates?*
- (15) *Doesn't John have yet to win the hearts of his classmates?*
- (16) *Oh, she has yet to finish, does she?*
- (17) *What do you have yet to eat?*

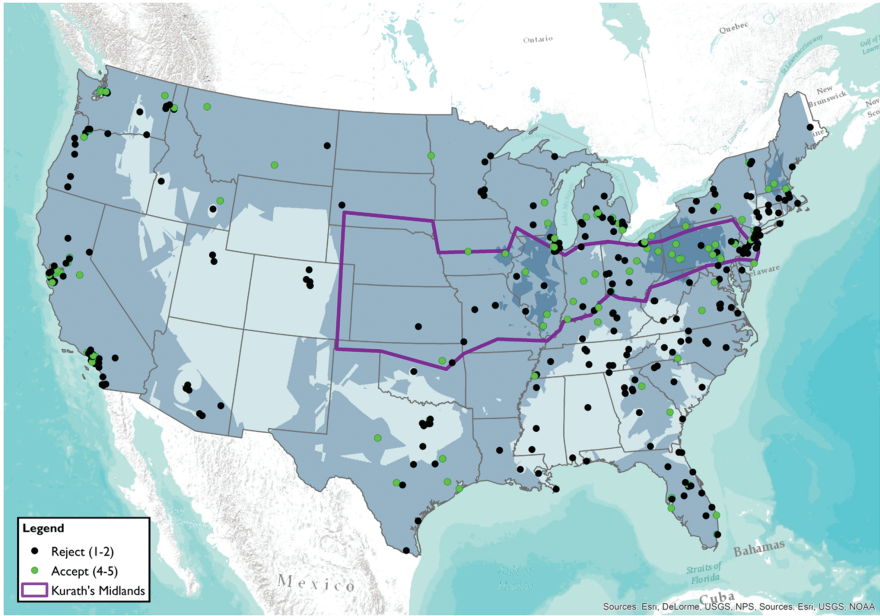
There turned out to be quite a bit of variation in general, such that the hot spots analysis did not pick out any region when applied to the sentences individually. However, this is likely due to there being a lot of variance in the responses, making a statistically significant signal hard to detect through the noise. When the judgments of (14)–(17) are averaged for each speaker, the hot spots analysis does find statistically significant hot and cold spots. These are shown in Map 17.



Map 17: Do-support with have yet to (average).

The hot spots picked out in this case do not in general seem to form neat isoglosses around the areas where many acceptances are found. For example, the strongest cluster of acceptances, indicated by the dark shading of the interpolation, is in western Pennsylvania, but that region does not fall into a hot spot region. The reason has to do once again with granularity. Outside of the area, there are many rejections, and they fall into the rather large critical distance determined by the tool to be optimal for this dataset.

It turns out that the areas picked out make sense from a dialectological perspective. The areas in question include Michigan, southern Wisconsin, and northern Illinois, Indiana, and Ohio. This area corresponds very closely to the area known in dialectology as the “Midlands,” more specifically, the “Upper Midlands.” Map 18 shows the region that Kurath (1949) identified tentatively as the Midlands. As discussed by Murray and Simon (2006), this region has historically been controversial, in that not all dialectologists agree that it should be considered a region in its own right. Carver (1987), for example, was generally interested in dialect region layering, and proposed that the area identified as the Midlands should more properly be considered a subregion of the North. Notwithstanding these disagreements, there seems to be a broad agreement that the region outlined in Map 18 is linguistically significant.



Map 18: Do-support with have yet to (average) with Upper Midlands region.

Supporting this conclusion, we find significant differences if we treat the different regions as separate populations, and compare the results. Thusly, the eastern half of the United States can be divided up into the Midlands (as illustrated in Map 18), the area to the north of the Midlands and the area to the south of the Midlands. When these three regions are compared in a two-way ANOVA, we find that the effect of region is significant ($F(3, 2560) = 17.23$, $p < 0.0001$). In terms of pairwise comparisons, only the difference between the South and the Midlands is significant. The difference between the North and the South is not significant, and the difference between the North and the Midlands is not significant either. This might be taken as support for Carver's (1987) contention that the Upper Midlands should be considered part of the North, although one should not be too hasty in such a conclusion. What is apparently undeniable is that the southern border of the Upper Midlands seems to be quite strong with respect to this one linguistic feature.

The results here show another way in which the hot spots analysis is useful. It is worth keeping in mind that significant results were only found once the judgments of four separate sentences (all testing for the same linguistic property) were averaged over. Even when this was done, the significant results did not always pick out neat and obvious isogloss regions. But they did tell us that the geographical distribution of the judgments was no accident. Moreover, when combined with existing knowledge about dialect regions, the results make sense: the hot spots analysis tells us that the geographical distribution is significant, but it is our existing knowledge of dialect variation that allows us to understand this significance. It is in the Upper Midlands that *do*-support is accepted most, and in the South that it is rejected most. Other areas exhibit varying degrees of variation.

Thus, we can tentatively say that *do*-support with HYT might be a genuinely Upper Midlands phenomenon. This may be so even if it is neither restricted to that area, nor found among all speakers there. In fact, as mentioned earlier, this is the case with many, if not most features of dialect variation, when studied carefully enough. It is quite rare to find a phenomenon that is truly restricted to a particular area, in the sense that it occurs among most speakers inside the region and among almost no speakers outside of the region. More frequently what we find is that it is a matter of degree: a phenomenon is found to a greater degree in one area than in the others. It is possible that *do*-support with HYT is one such case.

We will not dwell on the question here, however, because what is important presently is the role that the hot spots analysis plays in drawing such conclusions. More than in any of the other cases, the role of the hot spots analysis is somewhat indirect. There is a lot of variance in the judgments of *do*-support with HYT. We reduce the variance by averaging over four sentences. Then, the hot spots analysis helps to guide us toward a possible answer to the question of

whether there is geographically meaningful variation. The answer it provides is “yes” – the geographical variation is meaningful. What the meaning of it is requires further analysis. The hot spots analysis is thus only one of a broader set of tools we can use to analyze geographical variation in acceptability judgments. Nevertheless, we think it is a genuinely useful tool in that enterprise.

5 Conclusions

There are pros and cons to the hot spots analysis described in this paper. On the one hand, it picks out fine-grained patterns that are not easily visible at first glance. This, along with the fact that it replicates previous understanding about dialect geography quite well, makes it a very useful tool for geolinguistic research. On the other hand, it does not necessarily find all geographically relevant variation, and it does not necessarily find the “outer borders” of a phenomenon on its own. That is, it cannot be used as a blind, quantitative isogloss-drawing machine. Quite simply, it can tell us that geography matters for the acceptability of a given phenomenon, and provide us with clues as to where to look more closely. Even better, it can increase our confidence that the spatial dimension is relevant to a phenomenon without oversimplifying the facts. We can appreciate that there is geographic variation without trying to smooth over that variation too much – that is, without forgetting that there is often speaker variation inside and outside of the relevant region. Dialect regions are often a matter of the degree to which a phenomenon is sensitive to geography, rather than matter of a phenomenon existing within, and only within, a particular region. The hope is that future studies of syntactic acceptability judgments can make use of this tool, and others like it. By combining insights from dialectology and syntactic theory, we can cast a wider net, and gain an ever-increasing understanding of the subtleties of natural language syntax, and linguistic variation more generally.

Appendix A: Survey instructions and sample test sentence

Informal, casual language can be different in different places. The goal of this survey is to find out about your language, and the language spoken where you live and where you grew up.

We are not interested in what is correct or proper English

We are instead interested in what you consider to be an acceptable sentence in informal contexts. You will be presented with a sentence, or with a context plus a sentence. You will then judge the acceptability of that sentence on a scale of 1–5, with 1 being unacceptable and 5 being acceptable.

Here's you a piece of pizza

- ☐ 1 – totally unacceptable sentence, even in informal settings
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – totally acceptable
- ☐ Any comments?

It may help to read each sentence aloud before giving your judgment.

Appendix B: Map Metadata

Map 1 meta-information: Data comes from Survey 9, sentence 1116. There are 553 total data points, including 3s. Number of participants per judgment: 1 = 191, 2 = 124, 3 = 92, 4 = 62, 5 = 64. Hot spots was run with Optimized Hot Spots tool, on point data. 533 valid features, with a mean = 2.41 and SD = 1.383. 12 outlier locations. Optimal Fixed Distance Band determined with spatial distribution of features (average distance to 26 nearest neighbors) to be 319,541 meters. 273 output features statistically significant based on FDR correction.

Map 2 meta-information: Data comes from Surveys 6 and 8 combined, sentence 1094. There are 881 total data points, including 3s. Number of participants per judgment: 1 = 620, 2 = 154, 3 = 43, 4 = 26, 5 = 38. Hot spots was run with Optimized Hot Spots tool, on point data. 881 valid features, with a mean = 1.53 and SD = 1.02. 4 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 501,772.2444 meters. 149 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 0.029275368

Map 3 meta-information: Top: same as Map 2. Bottom: Same as Map 2, except hot spot analysis, where Fixed Distance Band was set to 150 km. 52 output features statistically significant based on FDR correction.

Map 4 meta-information: Data comes from Survey 8, sentence 1040. There are 520 total data points, including 3s. Number of participants per judgment: 1 = 148, 2 = 163, 3 = 106, 4 = 58, 5 = 45. Hot spots was run with Optimized Hot Spots tool, on point data. 520 valid features, with a mean = 2.40 and SD = 1.25. 2 outlier locations. Optimal Fixed Distance Band determined with spatial distribution of features (average distance to 26 nearest neighbors) to be 331,679 meters. 0 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 0.029275368

Map 5 meta-information: Data comes from Surveys 3–6, sentence 1043. There are 778 total data points, including 3s. Number of participants per judgment: 1 = 394, 2 = 172, 3 = 65, 4 = 29, 5 = 118. Hot spots was run with Optimized Hot Spots tool, on point data. 778 valid features, with a mean = 2.11 and SD = 1.45. 4 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 395,289.7458 meters. 303 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1464.48536

Map 6 meta-information: Same as Map 5.

Map 7 meta-information: Data comes from Surveys 3–6, sentence 1043. Including are only the points in the hot spot region from Maps 4/5. There are 197 total data points, including 3s. Number of participants per judgment: 1 = 72, 2 = 32, 3 = 22, 4 = 14, 5 = 57. Hot spots was run with Optimized Hot Spots tool, on point data. 197 valid features, with a mean = 2.7 and SD = 1.67. 3 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 183,683.168 meters. 48 output features statistically significant based on FDR correction. Interpolation same as Maps 4/5. It is worth noting that when the fixed distance was manually set to 100,000 meters, most of the green dot area in New England showed up as statistically significant hot spots.

Map 8 meta-information: Data comes from Survey 6, sentence 1086. There are 361 total data points, including 3s. Number of participants per judgment: 1 = 71, 2 = 86, 3 = 88, 4 = 71, 5 = 45. Hot spots was run with Optimized Hot Spots tool, on point data. 361 valid features, with a mean = 2.81 and SD = 1.30. 11 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 425,305.7511 meters. 40 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1267.96217476

Map 9 meta-information: Same as Map 8.

Map 10 meta-information: Same as Map 8, except that only 1s and 5s are represented in the point data.

Map 11 meta-information: Same as Map 10.

Map 12 meta-information: Data comes from Survey 6, sentence 1088. There are 361 total data points, including 3s. Number of participants per judgment: 1 = 26, 2 = 33, 3 = 57, 4 = 97, 5 = 148. Hot spots was run with Optimized Hot Spots tool, on point data. 361 valid features, with a mean = 3.85 and SD = 1.25. 11 outlier locations. Optimal Fixed Distance Band determined with spatial distribution of features (average distance to 18 nearest neighbors) to be 398,340 meters. 38 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1267.96217476

Map 13 meta-information: Data comes from Survey 6, sentence 1090. There are 361 total data points, including 3s. Number of participants per judgment: 1 = 43, 2 = 55, 3 = 69, 4 = 87, 5 = 107. Hot spots was run with Optimized Hot Spots tool, on point data. 361 valid features, with a mean = 3.44 and SD = 1.36. 11 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 293,349 meters. 35 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1267.96217476

Map 14 meta-information: Data comes from Survey 6, sentence 1092. There are 361 total data points, including 3s. Number of participants per judgment: 1 = 131, 2 = 84, 3 = 72, 4 = 45, 5 = 29. Hot spots was run with Optimized Hot Spots tool, on point data. 361 valid features, with a mean = 2.33 and SD = 1.30. 11 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 469,291.3013 meters. 31 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1267.96217476

Map 15 meta-information: Data comes from Survey 6, sentence 1091. There are 361 total data points, including 3s. Number of participants per judgment: 1 = 168, 2 = 98, 3 = 58, 4 = 19, 5 = 18. Hot spots was run with Optimized Hot Spots tool, on point data. 361 valid features, with a mean = 1.95 and SD = 1.13. 11 outlier locations. Optimal Fixed Distance Band determined with intensity of clustering at increasing distances to be 249,363.5502 meters. 28 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1267.96217476

Map 16 meta-information: Data comes from Survey 6, sentence 1093. There are 361 total data points, including 3s. Number of participants per judgment: 1 = 53, 2 = 38, 3 = 61, 4 = 84, 5 = 125. Hot spots was run with Optimized Hot Spots tool, on point data. 361 valid features, with a mean = 3.53 and SD = 1.43. 11 outlier locations. Optimal Fixed Distance Band determined with spatial distribution of features (average distance to 18 nearest neighbors) to be 398,340 meters. 57 output features statistically significant based on FDR correction.

Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 1267.96217476

Map 17 meta-information: Data comes from Survey 8, averages of sentences 1148, 1151, 1153, 1155. There are 520 total data points, including 3s. Number of participants per judgment: 1 = 97, 2 = 171, 3 = 147, 4 = 85, 5 = 20. Hot spots was run with Optimized Hot Spots tool, on point data. 520 valid features, with a mean = 2.43 and SD = 1.02. 2 outlier locations. Optimal Fixed Distance Band determined with spatial distribution of features (average distance to 18 nearest neighbors) to be 545,629.7981 meters. 112 output features statistically significant based on FDR correction. Interpolation was conducted with IDW, Power = 0.5, Search Radius 12, Output cell size = 0.029275368

Map 18 meta-information: Same as Map 18, with Upper Midlands region indicated instead of hot spots region.

References

- Ackerman, Lauren Marie. 2015. *Influences on parsing ambiguity*. Evanston, IL: Northwestern University dissertation. <http://gradworks.umi.com/37/41/3741393.html> (accessed 12 July 2016).
- Barbiers, Sjef & Hans Bennis. 2007. The syntactic atlas of the Dutch dialects. *Nordlyd* 34(1). doi:10.7557/12.89. <http://septentrio.uit.no/index.php/nordlyd/article/view/89> (accessed 12 July 2016).
- Bart, Gabriela, Elvira Glaser, Pius Sibling & Robert Weibel. 2013. Analysis of Swiss German syntactic variants using spatial statistics. In Ernestina Carrilho, Catarina Magro & Xosé Álvarez (eds.), *Current approaches to limits and areas in dialectology*, 143–169. Newcastle, UK: Cambridge Scholars Publishing.
- Behrend, Tara S., David J. Sharek, Adam W. Meade & Eric N. Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior Research Methods* 43(3). 800–813. doi:10.3758/s13428-011-0081-0.
- Benincà, Paola & Cecilia Poletto. 2004. A case of do-support in Romance. *Natural Language & Linguistic Theory* 22(1). 51–94.
- Boboc, Wellesley. 2016. *To hell and back: A syntactic analysis of hell in dialects of American English*. New York, NY: New York University Senior Honors Thesis.
- Bucholtz, Mary. 2006. Word up: Social meanings of slang in California youth culture. In Jane Goodman and Leila Monaghan (eds.), *A cultural approach to interpersonal communication: Essential readings*, 243–267. Chichester: Wiley-Blackwell.
- Bucholtz, Mary, Nancy Bermudez, Victor Fung, Lisa Edwards & Rosalva Vargas. 2007. Hella Nor Cal or totally So Cal? The perceptual dialectology of California. *Journal of English Linguistics* 35(4). 325–352. doi:10.1177/0075424207307780.
- Bybel, Kali & Greg Johnson. 2014. The syntax of “have yet to.” Paper presented at the 81st Southeastern Conference on Linguistics (SECOL), Myrtle Beach, SC, 27–29 March.

- Carver, Craig. 1987. *American regional dialects: A word geography*. Ann Arbor, MI: University of Michigan Press.
- Cassidy, Frederic G. & Joan Houston Hall. 2013. *Dictionary of American regional English*. <http://www.daredictionary.com> (accessed 12 July 2016).
- Champollion, Lucas, Ivano Ciardelli & Linmin Zhang. 2016. Breaking de Morgan's law in counterfactual antecedents. *Proceeding of SALT* 26. 304–324.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, CA: Sage.
- D'Alessandro, Roberta & Tobias Scheer. 2015. Modular PIC. *Linguistic Inquiry* 46(4). 593–624. doi:10.1162/LING_a_00195.
- Erlewine, Michael Yoshitaka & Hadas Kotek. 2016. A streamlined approach to online linguistic surveys. *Natural Language & Linguistic Theory* 34(2). 481–495. doi:10.1007/s11049-015-9305-9.
- Fetters, Michael & Aaron Steven White. 2016. Pseudogapping does not involve heavy shift. *West Coast Conference on Formal Linguistics (WCCFL)* 34. 205–213. <http://www.lingref.com/cpp/wccfl/34/index.html>.
- Fruehwald, Josef & Neil Myler. 2015. *I'm done my homework*: Case assignment in a stative passive. *Linguistic Variation* 15(2). 141–128. doi:10.1075/lv.15.2.01fru.
- Gibson, Edward, Steve Piantadosi & Kristina Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language & Linguistics Compass* 5(8). 509–524. doi:10.1111/j.1749-818X.2011.00295.x.
- Gibson, Edward & James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14(3). 225–248. doi:10.1080/016909699386293.
- Gordon, Peter C & Randall Hendrick. 1997. Intuitive knowledge of linguistic co-reference. *Cognition* 62(3). 325–370. doi:10.1016/S0010-0277(96)00788-3.
- Grieve, Jack. 2009. *A corpus-based regional dialect survey of grammatical variation in written Standard American English*. Flagstaff, AZ: Northern Arizona University dissertation.
- Grieve, Jack. 2014. A comparison of statistical methods for the aggregation of regional linguistic variation. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*, 53–88. Berlin & Boston: Walter de Gruyter.
- Grieve, Jack. 2016. *Regional variation in written American English*. (Studies in English Language). Cambridge: Cambridge University Press.
- Grieve, Jack, Dirk Speelman & Dirk Geeraerts. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change* 23(2). 193–221. doi:10.1017/S095439451100007X.
- Grieve, Jack, Dirk Speelman & Dirk Geeraerts. 2013. A multivariate spatial analysis of vowel formants in American English. *Journal of Linguistic Geography* 1. 193–221.
- Harris, James & Morris Halle. 2005. Unexpected plural inflections in Spanish: Reduplication and metathesis. *Linguistic Inquiry* 36(2). 195–222.
- Hartman, Jeremy. 2011. (Non-)intervention in A-movement: Some cross-constructional and cross-linguistic considerations. *Linguistic Variation* 11(2). 121–148. doi:10.1075/lv.11.2.01har.
- Harves, Stephanie & Neil Myler. 2014. Licensing NPIs and Licensing Silence: *Have/Be Yet To* in English. *Lingua* 148. 213–239. <https://doi.org/10.1016/j.lingua.2014.05.012>.

- Ipeirotis, Panos. 2010. Demographics of Mechanical Turk. Unpublished manuscript. New York: New York University. <http://crowdsourcing-class.org/readings/downloads/platform/demographics-of-mturk.pdf>, (accessed 24 October 2019).
- Irwin, Patricia. 2014. SO [TOTALLY] speaker-oriented: An analysis of “Drama SO”. In Raffaella Zanuttini & Laurence R. Horn (eds.), *Micro-syntactic variation in North American English* (Oxford Studies in Comparative Syntax), 29–70. Oxford: Oxford University Press.
- Karttunen, Lauri. 2014. Three ways of not being lucky. Paper presented at SALT 24, New York University. <https://web.stanford.edu/~laurik/presentations/LuckyAtSALTwithNotes.pdf> (accessed 30 May 2014).
- Kayne, Richard S. 2005. Some notes on comparative syntax: With special reference to English and French. In Richard Kayne (ed.), *Movement and silence* (Oxford Studies in Comparative Syntax), 277–333. Oxford: Oxford University Press.
- Kayne, Richard S. 2010. *Comparisons and contrasts*. Oxford: Oxford University Press.
- Kendall, Tyler & Valerie Fridland. 2016. Mapping the perception of linguistic form: Dialectometry with perception data. In Marie-Hélène Côté, Remco Knooihuizen & John Nerbonne (eds.), *The future of dialects*, 173–194. Berlin: Language Science Press.
- Ko, Edwin. 2016. *A corpus-based study of variation and change in adverb placement across world Englishes*. Washington, DC: Georgetown University Master’s thesis.
- Kotek, Hadas, Yasutada Sudo, Edwin Howard & Martin Hackl. 2011. Most meanings are superlative. In Jeffrey Runner (ed.), *Experiments at the interfaces* (Syntax and Semantics 37), 101–145. Leiden: Brill.
- Kurath, Hans. 1949. *A word geography of the eastern United States*. Ann Arbor, MI: University of Michigan Press.
- Murray, Thomas E. & Beth Lee Simon. 2006. What is dialect? Revisiting the Midland. In Thomas E. Murray & Beth Lee Simon (eds.), *Language variation and change in the American Midland: A new look at “Heartland” English* (Varieties of English Around the World G 36), 1–30. Amsterdam & Philadelphia: John Benjamins.
- Ord, J. Keith & Arthur Getis. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27. 286–306.
- Poletto, Cecilia & Paola Benincà. 2007. The ASIS enterprise: A view on the construction of a syntactic atlas for the Northern Italian dialects. *Nordlyd* 34(1). doi:10.7557/12.88. <http://septentrio.uit.no/index.php/nordlyd/article/view/88> (accessed 12 July 2016).
- Ravindranath, Maya. 2011. A wicked good reason to study intensifiers in New Hampshire. Paper presented at New Ways of Analyzing Variation 40, Georgetown University, October 27–30.
- Russ, Robert Brice. 2013. *Examining regional variation through online geotagged corpora*. Columbus, OH: The Ohio State University Master’s thesis. https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1385420187 (accessed 12 July 2016).
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.
- Sibler, Pius, Robert Weibel, Elvira Glaser & Gabriela Bart. 2012. Cartographic visualization in support of dialectology. Proceedings Auto Carto 2012, Columbus, OH, 16–18 September 2012. https://pdfs.semanticscholar.org/8ec3/6c6abf0ffd44e4233896f2252280bf2584d5.pdf?_ga=2.26013373.162660927.1570138439-1161082770.1567171093.
- Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43(1). 155–167. doi:10.3758/s13428-010-0039-7.

- Stoeckle, Philipp. 2016. Horizontal and vertical variation in Swiss German morphosyntax. In Marie-Hélène Côté, Remco Knooihuizen & John Nerbonne (eds.), *The future of dialects*, 195–215. Berlin: Language Science Press.
- Tamminga, Meredith. 2013. Phonology and morphology in Dutch indefinite determiner syncretism: Spatial and quantitative perspectives. *Journal of Linguistic Geography* 1(2). 115–124. doi:10.1017/jlg.2013.10.
- Torrego, Esther. 2010. Variability in the case patterns of causative formation in Romance and its implications. *Linguistic Inquiry* 41(3). 445–470. doi:10.1162/LING_a_00004.
- Tyler, Matthew & Jim Wood. 2019. Microvariation in the *have yet to* construction. *Linguistic Variation Online First*. <https://doi.org/10.1075/lv.16006.tyl>.
- Wood, Jim. 2014. Affirmative semantics with negative morphosyntax: Negative exclamatives and the New England *So AUXn't NP/DP* construction. In Raffaella Zanuttini & Laurence R. Horn (eds.), *Micro-syntactic variation in North American English* (Oxford Studies in Comparative Syntax), 71–114. Oxford: Oxford University Press.
- Wood, Jim, Laurence Horn, Raffaella Zanuttini & Luke Lindemann. 2015. The Southern dative presentative meets Mechanical Turk. *American Speech* 90(3). 291–320. doi:10.1215/00031283-3324487.
- Wood, Jim, Raffaella Zanuttini, Laurence Horn & Jason Zentz. 2019. Dative Country: Markedness and geographical variation in Southern dative constructions. *American Speech (Advance Publication)* doi:doi.org/10.1215/00031283-7587901.
- Yerastov, Yuri. 2010. *I'm done dinner: When synchrony meets diachrony*. Calgary, AB: University of Calgary dissertation.
- Zanuttini, Raffaella & Paul Portner. 2003. Exclamative clauses: At the syntax-semantics interface. *Language* 79(1). 39–81.
- Zanuttini, Raffaella, Jim Wood, Jason Zentz & Laurence R. Horn. 2018. The Yale grammatical diversity project: Morphosyntactic variation in North American English. *Linguistics Vanguard* 4(1). 1–15.