NEEDED RESEARCH IN AMERICAN DIALECTS:

VARIATION IN MORPHOSYNTAX

Jim Wood and Raffaella Zanuttini

Yale University

INTRODUCTION

The goal of the volume is to sum up research done in the past two decades and chart the way forward. The goal of this chapter is to provide an overview of the new sources of data, new methodologies, and new lines of research that have been developed in the past two decades focusing on the the study of morphosyntactic variation in American English. This overview will also serve to chart the way forward for research in this area, as it will highlight discoveries that are being made and questions that have emerged and are ripe for further investigation.

Since the pioneering work on linguistic variation by Labov (1966, 1972), much of the research on dialect variation in North America has centered on methodological approaches that require what has become known as a sociolinguistic variable. A sociolinguistic variable involves two or more variants that are essentially interchangeable in a given context. For example, the progressive *-ing* suffix can be pronounced as [ɪŋ] or [ɪn], and the choice has no semantic consequences. Researchers determine the envelope of variation—the contexts where all the variants can occur in principle—and use statistical methods to study how different linguistic and non-linguistic (e.g. social) factors condition the choice of one variant over another in a corpus of spoken or written language. This method has been used productively to study numerous syntactic constructions, including expressions of stative possession (*have*, *have got*) (Tagliamonte et al. 2010), necessity modals (*must*, *have to*, *have got to*, etc.) (Tagliamonte and D'Arcy 2007), intensifiers (*so*, *really*, *very*, etc.) (Tagliamonte and Roberts 2005; Tagliamonte 2008), quotative constructions (D'Arcy 2004; Tagliamonte and D'Arcy 2004), comparative adjectives (*more angry*, *angrier*) (D'Arcy 2014), complementizer deletion (*I think that/Ø she. . .*) (Blondeau and Nagy 2008), and particle shift (*pack the car up*, *pack up the car*) (Röthlisberger and Tagliamonte 2020).

Applying the notion of a sociolinguistic variable to syntax, however, can in many cases be challenging. First, many alternations in syntax do have semantic consequences, and this alone has been seen as a major stumbling block for researchers attempting to study morphosyntactic variation as an alternation between two or more variants (see discussion and references in Bresnan and Ford 2010, 170). In some cases, it can be argued that supposed meaning differences are not categorical or even non-existent, or that the envelope of variation can be defined in such a way to minimize the effect of the supposed differences (Bresnan et al. 2007; Kendall et al. 2011; see also Singler 2001). Nevertheless, there remain many cases, including many of the constructions discussed in this chapter, where the choice of a syntactic variant has strong enough semantic consequences to warrant studying that variant as a separate syntactic construction.

Second, many dialect-specific syntactic constructions do not clearly alternate with anything else. Hasty (2014) refers to these as 'Type 2' syntactic variables, and argues that multiple modal constructions are like this: sentences like *You might should eat before you go* do not have any obvious 'alternative form' that they alternate with. There are numerous constructions like this (see Hasty 2014, 272), although in many cases, whether a construction is Type 2 or not may depend on how strict we are with the semantic equivalence criterion discussed above. If the choice between Variant A and Variant B has a clear semantic effect, then it might be fair to say that Variant A is actually a Type 2 variable, because it has no alternate form that preserves its meaning.

Third, many syntactic constructions do not occur frequently enough to study in corpora. As Robinson and Duncan (2019) point out:

> "The difficulties of such rarity extend beyond the need for robust results; the infrequent occurrence of syntactic variables also poses a challenge to establishing the envelope of variation. After all, when tokens of the relevant construction are rare, the presence of one variant and absence of another in a given environment does not necessarily entail that one variant must (or cannot) be used in that environment. (Robinson and Duncan 2019, 210)

To make matters even more challenging, the envelope of variation itself might even change over

time (Brook 2018).

Finally, there are aspects of dialectally restricted constructions that are worthy of study that go beyond the factors that condition the way that they alternate with other variants (or constructions). On the one hand, the geographic and non-geographic social factors that affect the acceptability and/or production of these constructions can be studied independently of other variants while still shedding light on the landscape of linguistic variation. On the other hand, the syntactic properties of these constructions can be studied independently of other variants and deepen our understanding of natural language syntax in general. In the remainder of this chapter, we will discuss both of these aspects of dialect syntax in more detail, as well as new resources for data and analysis that can be used to build on the insights from previous research on syntactic variation, and perhaps even offer new ways to approach some of the challenges that arise in research on syntactic variation.

We will refer to a number of different syntactic constructions and phenomena throughout this chapter, whenever they illustrate a general point we are making. Rather than introduce them all as we go, repeatedly interrupting the main point of the discussion, we introduce them here along with the names that we use to refer to them:

(1)   I would have rathered {go/gone} to a small school.   (Verbal *rather*)

(2)   Do you want to come with?   (The *come with* construction)

(3)   The car needs washed.   (The *needs washed* construction)

(4)   Are you done your homework?   (The *be done my homework* construction)

(5)   Do you have yet to visit your grandma?   (The *have yet to* construction
                                                                          with *do*-support)

(6)   Here's you a piece of pizza.   (Dative Presentatives)

(7)   He plays guitar, but so don't I.   (The *so don't I* construction)
        ≈ '…so do I'

(8)   Football is more popular than baseball anymore.   (Nonpolarity *anymore*)

(9)   This seat reclines hella.   (Intensifier *hella*)

≈ 'This seat reclines a lot.'

(10)   I need me some black jeans.   (Personal Datives)

We cannot discuss these constructions in any detail here, but see the Yale Grammatical Diversity Project website for overviews of these and other instances of morpho-syntactic variation found in North American English.

NEW SOURCES OF DATA AND METHODS OF ANALYSIS

In the past few decades, numerous new sources of data and methods of analysis have been developed, and continue to be developed to study morphosyntactic variation. In many cases, we have only begun to make use of these resources. In this section, we discuss the proliferation of corpora, the availability of online surveys, and the possibility of developing new webtools and interactive dashboards for engaging with morphosyntactic data.

PROLIFERATION OF CORPORA. The general development of technology in the past few decades has of course been staggering. Within linguistics, this has been paralleled by developments in corpora in at least two ways. First, specifically curated online corpora, in many cases tagged or even annotated syntactically, have been and continue to be developed. For example, the currently available version of the *Corpus of Contemporary American English* (COCA) is far more expansive and sophisticated than the version that was available 10 or 15 years ago (and the fact that it is available in the first place is a major advancement for the field). The same can be said for other corpora, databases and online resources, including *The Audio-Aligned and Parsed Corpus of Appalachian English* (AAPCAppE) (Tortora et al. 2017), the *MultiMo* database, and the eWAVE Atlas (Kortmann and Lunkenheimer 2013), among others. Resources like these provide more opportunities to study morphosyntactic variation (recent examples include Tortora 2018; de Clercq and Vanden Wyngaerd 2019; Blanchette and Collins 2019; Van Gelderen 2019; Merchant 2021; Donati et al. 2021).

In addition to these specifically curated corpora, there is now an enormous amount of online language data that is available for study, ranging from simple Google searches to geotagged Twitter studies. Linguists have used data from Google searches to argue that previous empirical generalizations were incorrect (Bruening 2014, 2018; Harley and Jung 2015) or simply to provide more naturalistic examples of non-standard constructions (Wood 2013, 2014; Tyler and Wood 2019). Numerous Twitter studies have emerged studying dialect variation (Jones 2015, 2016; Huang et al. 2016; Robinson and Duncan 2019; Strelluf 2020; Willis 2020). These methods are useful for constructions that can easily be searched using exact strings or morphological forms. For example, it is quite simple to search for verbal *rather* by looking for the form *rathered*. However, these methods are not as useful for constructions that are not easily searched for in this way. For example, it is harder to search for examples of the *come with* construction (as in *Do you want to come with?*) and exclude sentences where *with* takes an overt complement (as in *Do you want to come with us?*).

SURVEYS AND CROWDSOURCING. Surveys have long been used to study morphosyntactic variation, but it is now easier than ever to administer, collect and analyze survey data. In particular, the emergence of crowdsourcing platforms such as Amazon Mechanical Turk, Crowd Flower, and Survey Monkey allow surveys to be distributed over an enormous geographical region (in principle the whole world) in a very short time. The responses are of course received digitally, so processing and analyzing the surveys can be much quicker. Sprouse (2011) originally showed that data collected in this way was at least as reliable as data collected in person, and later work showed that dialect variation could be meaningfully collected and analyzed with such platforms as well (Wood et al. 2015, 2020b; Kim et al. 2019; Wood 2019). The advantage to such platforms is that they allow the researcher to choose exactly which questions to ask, and rare or complex syntactic constructions that are not as easily searchable in corpora can be designed and tested. Internet searches can still be useful; as pointed out by Wood et al. (2020a), the internet can be used to find natural-sounding examples which can then be included in surveys. In addition,

however, it is also easier than ever to use social media platforms such as Facebook and Reddit to recruit survey participants.

NEW WEBTOOLS AND INTERACTIVE MAPS. It is now possible to make the results of research available to a wider audience by developing webtools such as dashboard and interactive maps. For example, the Scots Dialect Atlas online has a version that is intended for a lay audience and a version that is intended for linguists. Similar tools have been developed by the Yale Grammatical Diversity Project (henceforth, YGDP) for mapping their survey data, either in part or in a more comprehensive manner. Users can search for various morphosyntactic phenomena and generate maps of their searches, which can be useful for the syntactic analysis of such phenomena (Thoms et al. 2019; Wood 2022).

MAPPING SOFTWARE. Numerous new kinds of software and statistical packages have come into existence in ways that are much more accessible to linguists. One notable mention is Gapmap (Leinonen et al. 2016), which offers a fairly wide range of analytic possibilities, many within the general approach to dialect mapping known as dialectometry (Wieling and Nerbonne 2015), and is fairly easy to use even for relative beginners. Here, however, we focus on Geographical Information Systems (GIS) Software, such as ArcGIS, QGIS, and the GIS packages that are available in R (such as ShinyDialect, described in Chagnaud et al. 2017). Such software can have a somewhat steep learning curve, but certainly can be learned by a linguist who is willing to. The kinds of geospatial statistical analysis that might be useful to linguists have only begun to be explored (see Grieve 2016), and many of the methods that have been explored have not yet been applied to morphosyntactic data, even though they easily could be. Even the ways that linguistic data could be visualized spatially have only begun to be explored (Sibler et al. 2012; Stoeckle 2016).

To give one kind of example, in the YGDP surveys, we collected acceptability judgments on a scale of 1–5. In most of our work, we have analyzed the survey responses geographically by treating the judgment scores as continuous, and plotting every speaker in their primary childhood

location. Statistical tools like the G_i^* "hot spot" analysis and interpolation are then based on these data under the assumption that they are continuous—that the difference between a 2 and 3, for example, is equal to the difference between a 3 and 4, so a judgment of 2 and 4 could be averaged together as a judgment of 3. However, one could equally well justify dividing speakers into two qualitative categories, "accept" and "reject", by choosing a cutoff point such as "4 or above". After all, it is well known that speakers do not in practice treat Likert scales as continuous (Sprouse 2007),[1] and cutoff points are often chosen in work on dialect variation to turn continuous data into discrete data (Labov et al. 2006). Then speakers from certain regions could be collapsed together into a data point along the lines of "percentage of speakers who accept X". Or, if there are multiple sentences of a particular construction, they could be collapsed together as "percentage of sentences in construction X that speaker Y accepts".

To be more concrete, suppose we are evaluating speakers' judgments of the *needs washed* construction, and there are four *needs washed* sentences on a survey. Speaker A is unsure of the construction, and rates them all as 3. Speaker B accepts some sentences and not others, say, with ratings of 1, 2, 4, and 5. If these scores are averaged, both speakers end up with a 3. But if we instead say "percentage accepted" with a cutoff of 4, Speaker A gets a 0% and Speaker B gets 50%. In one case, we map out two 3s, in another, one 0 and one 50—very different empirical pictures.[2] What is needed in this area is the development of a set of best practices and guidelines for analyzing linguistic variation and for visualizing it. By this we do not mean that there should be a "one size fits all" kind of solution, but rather a better understanding of the different ways that data can be analyzed and visualized spatially and when each method is appropriate. Different techniques should be tried and compared in order to flesh out the basic range of possibilities and determine which options are the most useful.[3]

<center>GENERATIVE SYNTACTIC THEORY AND DIALECT VARIATION</center>

Over the years, generative syntactic theory has seen a steady growth of interest in dialect variation. Much of this interest has roots in Richard Kayne's early work comparing French and

<center>7</center>

English (Kayne 1969), a line of work that led to studies showing the theoretical benefits of comparing increasingly closely related languages and dialects (Kayne 1989, 1991, 1997). The development of the notion of the PARAMETER as a source of linguistic variation within the Government-Binding (GB) framework provided a clear way to generate hypotheses about how languages differ, and closely related languages and dialects played a major role in the formulation and exploration of parameters. A very fruitful line of research developed starting from the study of dialects spoken in Northern Italy (Benincà 1989, 1994; Poletto 1992, 1993, 1996; Zanuttini 1991, 1995, 1997) expanding to the variety of English spoken in Belfast (Henry 1992, 1995), to the varieties found in the Netherlands and Flanders (Barbiers et al. 2002, 2008, 2016), and to many more all over Europe and beyond.

The range of variation found across dialects offered precise challenges to parametric theory that led some to conclude that it was inadequate as a theory of variation (see discussion in Svenonius 2008), although Kayne (2005) has defended a version of this theory that ties parameters to function morphemes (an idea that goes back to Borer 1984), and his work is explicitly built on dialect variation. Roberts (2019) has argued for a parameter hierarchy that allows for large, macro-parameters in addition to micro-parameters that distinguish between closely related dialects (see also Baker 2001, 2008; Van Craenenbroeck et al. 2019). Adger (2006) has argued that the basic architecture of the Minimalist Program (Chomsky 1995, 2001) is particularly well-suited to handle dialect variation, and even intra-speaker variation (see also Adger and Smith 2005). Regardless of what position one took on parameters, dialect variation became an increasingly central and important part of generative theorizing, and the focus of numerous edited volumes and special issues of journals (Cornips and Corrigan 2005; Trousdale and Adger 2007; Zanuttini and Horn 2014; Etxepare and Gallego 2018–2020), and this interest was not restricted to mainstream generative theory (see Bender 2001 and various papers in the above-cited volumes); see also Duncan (2019, 2021) and Strelluf (2022) for excellent examples of empirically-detailed research that engage with syntactic theory in a serious way.

In this section, we discuss two particular ways in which the already fruitful relationship

between the study of dialect variation and morphosyntactic theory needs to be expanded upon. First, non-standard syntactic constructions benefit from the increasingly sophisticated toolbox provided by syntactic theory. Second, theoretical investigations often lead to the discovery of empirical points of variation that might not otherwise be noticed. We discuss each of these in turn.

SYNTACTIC ANALYSIS OF CONSTRUCTIONS.   There have been a number of cases where a certain syntactic phenomenon has been described by linguists who study varation, but who do not specialize in morphosyntactic analysis or morphosyntactic theory (dialectoligists, sociolinguists, phonologists, etc.). Examples include multiple modals, negative concord, personal datives, dative presentatives, and many others. An important task moving forward is to provide detailed, theoretically grounded analyses of these phenomena. Such analyses provide an important testing ground for the empirical adequacy of the theory, and very often special insight into the nature of language. Beyond that, they provide a finer-grained characterization of the nature of the phenomenon and often further variation within the phenomenon. For example, Fruehwald and Myler (2015) analyze the *be done my homework* construction, where speakers of some dialects accepts and produce sentences like *I am done my homework* (see also Yerastov 2015). This is often described as containing a silent preposition like *with* or a verb like *doing*, or it is described as a kind of perfect aspect with *be* instead of *have*. Fruehwald and Myler (2015), however, provide a battery of diagnostic tests to argue persuasively that it is better analyzed as an adjectival passive with the unusual property that it can assign accusative case to a complement. This is an interesting conclusion for the theory of adjective formation (Alexiadou et al. 2014; Bruening 2014), but also provides a new lens from a dialectal standpoint: Myler (2016) reports that since the publication of Fruehwald and Myler (2015), he has found some speakers whose judgments seem to show that they in fact have a silent *with*.

Variation of this kind is detectable, but one must know what questions to ask, and it is likely to fly under the radar within and across communities, from the perspective of both speakers and linguists. It is entirely possible, for example, that the apparent geographic distribution of a

construction like the *be done my homework* construction is misleading because people who say *I'm done my homework* do not in fact have the same construction. Another example of this comes from Tyler and Wood (2019) in their study of the *have yet to* construction. They point out that while most speakers have some version of it—most can say *I have yet to visit her* meaning 'I have not yet visited her'—speakers differ as to whether they treat *have* as a main verb or as an auxiliary verb (among other things). Wood (2019) shows that the availability of the "main verb" analysis is affected at least in part by geography, in that speakers from the Midland region are more likely to have this version of the construction. Here again, subjecting the construction to a fine-grained, theoretically driven syntactic analysis reveals a layer of linguistic variation that we would not notice if we simply asked where speakers accept the *have yet to* construction, without distinguishing whether they treat *have* as a main verb or as an auxiliary.

Another positive consequence of attempting to provide precise analyses of non-standard constructions is that we are often forced to ask questions that we might not have otherwise asked about related constructions found in more mainstream dialects. In analyzing the *so don't I* construction (Wood 2014), one discovers interesting puzzles about negative exclamatives and the *so do I* construction in mainstream dialects (Wood 2008). In researching the *be done my homework* construction, one finds that there is no clear or standard analysis of the more widespread construction *be done doing my homework* (Biggs 2021). In analyzing presentative dative sentences like *Here's you a piece of pizza* (Wood and Zanuttini 2018), one discovers numerous challenges in the analysis of more widespread presentative sentences like *Here's a piece of pizza* (Wood and Zanuttini 2023). Time and again, the analysis of non-standard syntactic constructions points us to more widespread linguistic puzzles that provide fertile ground for novel insights into language.

In sum, as we continue to discover and rediscover the existence of syntactic constructions that do not exist in all dialects, what is needed is detailed syntactic analysis of those constructions. There has certainly been some work in this area, and it seems to be growing (one can point to some of the references above, as well as others such as Wood 2013, Johnson 2014,

10

Myler 2013, and most of the papers in Zanuttini and Horn 2014), yet much more work remains to be done. There also needs to be more debate: we generally learn more when linguists with different perspectives rigorously pursue distinct analyses than we do when a construction is only analyzed once.

DISCOVERING VARIATION IN THE CONTEXT OF THEORY. When doing research in syntactic theory, it is quite common to encounter sentences that are accepted by some but not all speakers, without finding a social factor (such as age or geographical region) that correlates with such variation. Frequently, this is annotated with a "%" and, depending on the argument being developed, the fact that *some* speakers accept the sentence may be enough to make the argument go through. We have termed such cases "variation in every room" (Zanuttini et al. 2018; Wood et al. 2020a). One does not know until one looks, and it may be of substantial interest to verify empirically that a phenomenon really does vary at the level of the individual. In our experience, it is not uncommon for linguists to assume that any variation that they encounter can be traced to a particular, socially-defined dialect, and some linguists express skepticism that idiolectal variation truly exists.

The YGDP has, over the years, found a number of syntactic constructions whose acceptability cannot as of yet be tied to any social factor. These include (among others) those exemplified in the sentences below, which illustrate copy-raising from object position (11) (Landau 2011; Asudeh and Toivonen 2012); double-aux raising (12) (Johnson 1988); *feel like that* sentences (13) (Kaplan 2019; Srivastava 2020); promise-control (14) (Hartman 2011); and swiping without sluicing (Kayne 2015, 16) (15). All examples are included in Wood et al. (2020a):

(11)   John seems like Mary offended him. ('copy-raising from object position')

(12)   Shouldn't have Pam remembered her name? ('double-aux raising')

(13)   I feel like that we should win this game. ('*feel like that*' construction)

(14)   John threatened me to come to my house. ('promise-control')

(15)   What about were you talking? ('swiping without sluicing')

In at least one case, a point of variation that was initially discovered in the course of theoretical work turned out to have geographical correlates: recent research on the *have yet to* construction (exemplified in (5) above) was not initially concerned with variation, but instead pursued the analysis of the construction under the assumption that it would be the same for all speakers (Kelly 2008, 2012). However, subsequent research made it clear that speakers varied substantially as to whether they allowed *have* to be treated as a main verb or an auxiliary verb (Bybel and Johnson 2014; Harves and Myler 2014). Consequently, Tyler and Wood (2019) pursued a syntactic analysis that took this variation into account, and Wood (2019) conducted a statistical analysis of the geographic distribution of *do*-support (as a diagnostic for the "main verb" status of *have*), and showed that geographical region does indeed play an important role.

Moving forward, future studies of dialect variation in American English syntax may take into account the status of *have* in *have yet to*. More broadly, we suspect that a greater cross-fertilization of theoretically-driven work and studies of dialect variation will lead to more cases where the kind of variation discovered by theoretical syntacticians will be of interest to dialect studies.

INTERACTIONS BETWEEN GEOGRAPHY AND OTHER SOCIAL FACTORS

It is well known that linguistic variation may correlate with not only geographical region, which is the most common target of investigation for dialect studies, but other social categories as well, including age, gender, race, socio-economic status, and education. As discussed earlier, even capturing or understanding the nature of geographically-based dialect variation is no easy feat. In some cases, geography and other social categories are both relevant. For example, non-polarity *anymore* correlates with region, but also with age (older speakers find it more acceptable), the urban/rural distinction, and perhaps other social categories. However, the challenge is compounded by the fact that, in principle, geography and other social factors may interact. Suppose, for example, that a particular phenomenon is correlated with age in one area but not

12

another. At present, we do not know of any clear way to discover such cases, unless we already have a reason to suspect what the patterns might turn out to be.

In our own dataset, we know of some cases where the regional picture is quite different depending on which age group is examined. The use of *hella* is widely considered a Northern California-based feature. Indeed, Northern California is picked as a statistically significant hot spot when all age groups are included. However, when only younger speakers are included, the distribution of *hella* does not appear to be as strongly distinctive of that region. We find a similar situation with Personal Datives, which are widely considered to be a Southern construction. While that is true, among younger speakers the construction appears to be much more widespread, and not distinctive to the South. In both of these cases, it seems that a construction that was once characteristic of a particular geographical region has become at least familiar, and perhaps even acceptable, among younger speakers far beyond that region. Among older speakers, the regional distinctions still hold.

In these cases, the pattern we find shows that there is a statistically reliable geographic component to the distribution of a construction. But when we restrict it to a socially-defined subset of the population, the regionality disappears (e.g. among younger speakers) or gets stronger (e.g. among older speakers). Another case of the latter comes from the *needs washed* construction, where we find a distinctive geographical distribution: it is most robustly accepted in the Midland region. However, if we restrict our mapping to only participants who are classified by the census as living in a "rural" area, the geographical pattern goes from distinctive to nearly exceptionless.

These fairly simple cases show that geography and other social factors can interact with each other. In all these cases, the only reason we know is because we had reason to look—for example, because we saw from initial data that the interacting factors were independently important. But statistical interactions are often more complex than this, and can hide the "main effect". Suppose that, for some construction, the interaction of, say, geography and age is such that, when all ages are included, no geographical pattern is found at all, and no independent age effect, either. For

13

example, older speakers in rural areas in the North accept a construction, as do younger speakers in urban areas in the South, and in the West there is individual variation that is not tied to age or the urban/rural distinction. It is quite possible that, in this situation, we would not detect the effects of any of these factors, and that we would categorize it as a case of "variation in every room". This hypothetical example is artificial in a way that may make it seem far fetched, but it serves to make a real point: social factors really do interact with each other in complex ways, and if we don't know about those interactions in the first place, we don't know to look for them.

If we already know what the important dialect regions are, we can use that information as a starting point. For example, a simple statistical regression that includes dialect region as a factor can be set up to look for interactions. The problem is that dialect regions are gradient, and different regions are important for different phenomena. So in fact, we do not really know what the relevant dialect regions are, at least not necessarily. If, say, age is only important in one particular area, we would only be able to detect that if the area happened to coincide with the areas that we feed into the regression analysis to begin with.

If we know in advance what to look for, there are ways to proceed, and, indeed, recent research shows that interactions of this sort do occur, and are perhaps quite common. For example, Jones (2015) analyzed Twitter data to see if the geographical variation found in African American English (AAE) dialects lined up with geographical variation based on white speakers. He found that in fact, there is an entirely different set of dialect regions for AAE, which exists alongside the dialect regions discovered for white speakers.

The key here is that there was reason to look in the first place. If the dataset is large enough, one can take any social category, and draw a separate set of maps to discover the dialect areas relevant to that category. For example, we could make a whole dialect atlas, and use whatever aggregation methods we deemed appropriate, for speakers under the age of 30, and another one for speakers over 30. We could propose a whole set of maps based only on people who identify as male, and another based only on people who identify as female. In practice, however, this is not so easy. First, it requires a lot of data. Second, we still have to make arbitrary decisions about

what the cutoffs or categories are. Third, we still don't get at the overall interaction problem, because each category will be heterogeneous in other ways (people identifying as female will have a range of ages, races, etc.).

Even if we had enough data to drill down and create entire atlases for every conceivable combination, and we were confident about our choices, we do not have a reliable way to compare maps with each other statistically. How different do they have to be in order to be considered reliably distinct? And based on whatever that metric might be, how do we correct for multiple comparisons? We would be comparing hundreds of maps, so we would need to have some way to be confident that our results were not filled with false positives and false negatives.

In fact, the problem of comparing two maps is actually quite general, and solving this problem would be a major advancement for dialect studies, independently of the issue of interaction. To illustrate, consider the maps in Figures 1–4, taken from Wood et al. (2020a). The maps in Figures 1 and 2 are very similar, and in reality, probably reflect the same distribution. That is, they are not identical, but the differences between them are probably not meaningful. However, this is at present just a judgment call. In the case of the maps in Figures 1 and 2, it is perhaps a relatively safe judgment call; but what about the map in Figure 3? It is similar to the maps in Figures 1 and 2, but also clearly different. How different is it? Are the differences meaningful? Do we decide that it is a fundamentally distinct distribution? If we look at the map in Figure 4, we can be fairly confident that the distribution is fundamentally different: a different construction found in a distinct—even if overlapping—dialect region.
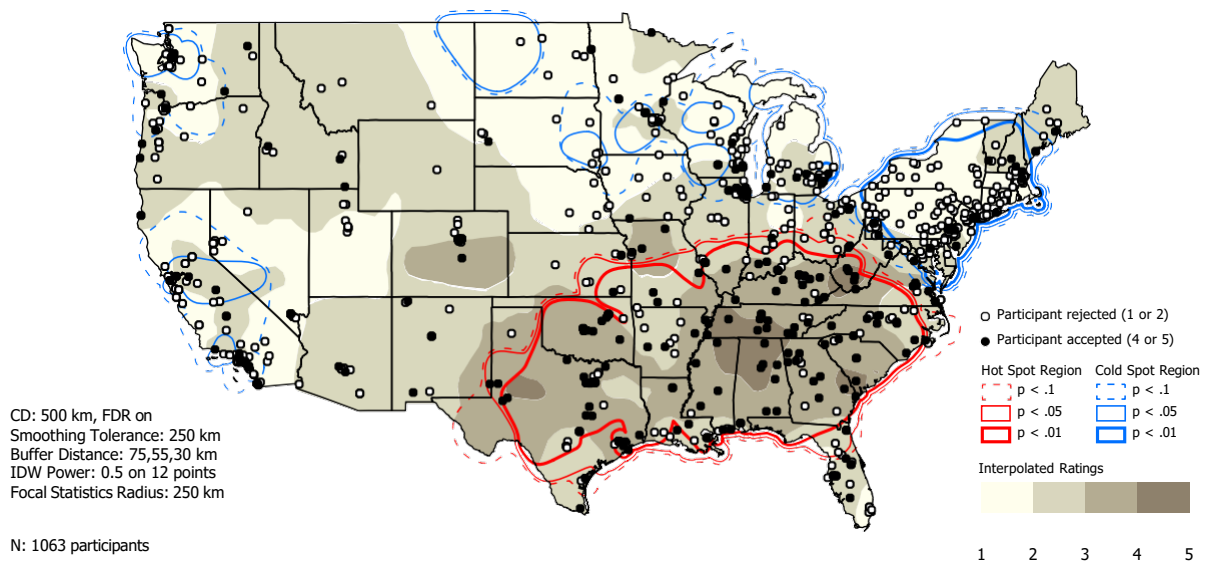
F1070: "Here's me a good pair of jeans!"

CD: 500 km, FDR on
Smoothing Tolerance: 250 km
Buffer Distance: 75,55,30 km
IDW Power: 0.5 on 12 points
Focal Statistics Radius: 250 km

N: 1063 participants

Figure 1 Here's me a good pair of jeans!

F1071: "Here's us a gas station—pull over!"

CD: 500 km, FDR on
Smoothing Tolerance: 250 km
Buffer Distance: 75,55,30 km
IDW Power: 0.5 on 12 points
Focal Statistics Radius: 250 km

N: 1602 participants
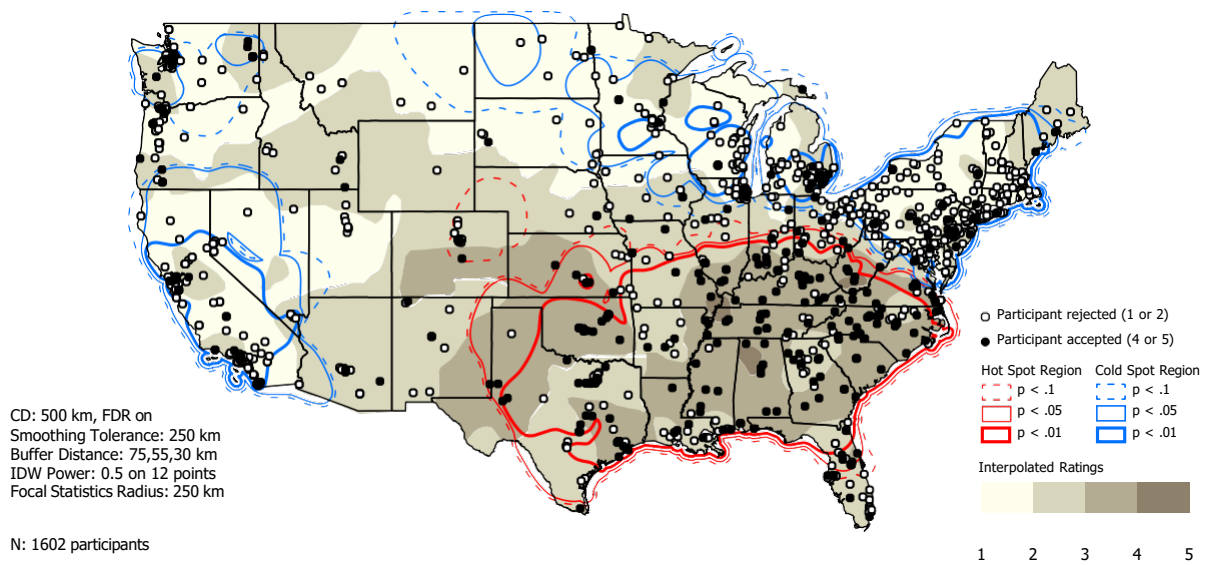
Figure 2 Here's us a gas station—pull over!

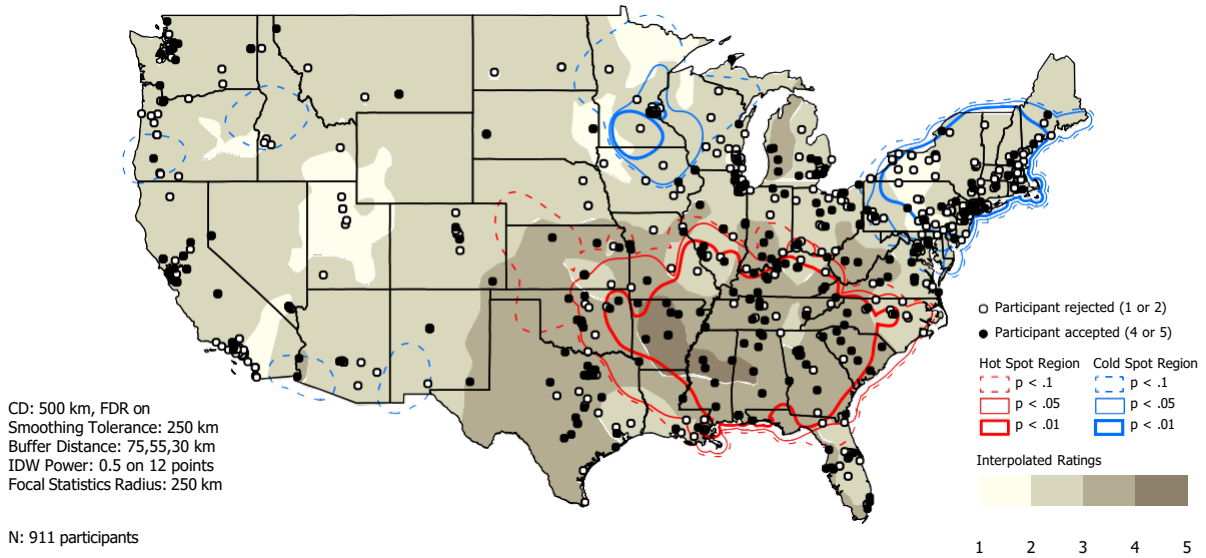F1180: "I'd put me a marble or two in my pocket."

Figure 3 I'd put me a marble or two in my pocket.
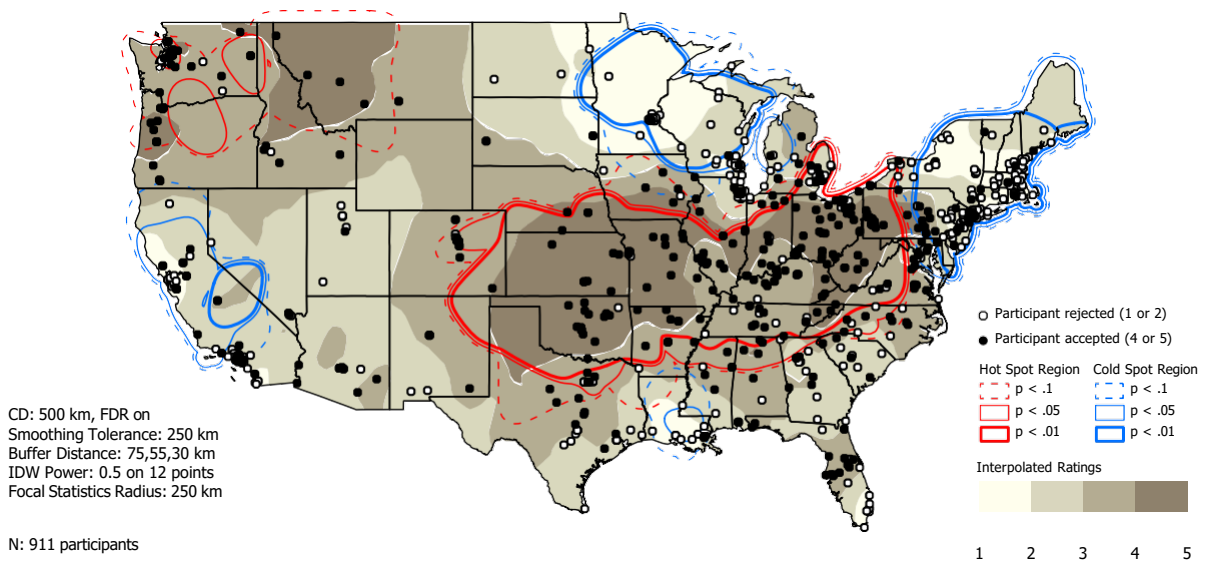
F1181: "My car needs fixed."

Figure 4 My car needs fixed.

Returning to other dimensions of social variation and their interaction with geography, what is needed is a way to detect geographically-based variation the way that, say, the G_i^* statistic

does, but in a way that can also be sensitive to other categories in each area. There has been some promising work using Geographically Weighted Regression (Willis 2017, 2019), but, so far, it has not been shown that it can be applied in a general way to a dataset to generate results in a bottom-up manner. For example, the methodology in Willis (2017, 2019) requires a particular kind of data, where what counts as a "place" is predetermined, and regressions are run in different places, weighted according to where they are. Expanding this work so as to provide a general way of looking for interactions between geography and other social categories will pave the way for a whole host of new discoveries that can be pursued in follow-up case studies.[4]

CONCLUSION

The past few decades have been a particularly fertile time for the study of syntactic dialect variation, and the decades ahead look, if anything, even more promising. Linguists have at their disposal more tools and resources than ever, and continue to develop and use them in novel and creative ways. Formal theories offer precise ways to ask specific questions about dialect variation, and theoretical syntacticians are increasingly taking up such questions. We expect that in the future, there will be more intellectual cross-fertilization between dialectologists, sociolinguists, and syntacticians, resulting in more resources, deeper analysis, and a more comprehensive empirical picture of syntactic dialect variation than we have ever had before.

NOTES

1. For discussion and additional references, see Tyler and Wood (2019, 236–237).
2. One might try to incorporate a measure of deviation such as the Standard Deviation into the final score, but it is not clear how to incorporate this information into a map. One possibility would be to explicitly map out the amount of variation, rather than the value itself; see Stoeckle (2016) for a study that does something like this.
3. For example, it is now quite straightforward in GIS software to create a map that visualizes geographic variation in acceptability judgments as though it were a mountain range, with the

peaks being the highest acceptance (or rejection) and the valleys the opposite. However, despite the fact that this can create striking images, we have not yet found it to be at all useful.

4. Another dimension of variation that we have not discussed involves meta-linguistic awareness. To what extent does the presence or absence of stigmatization of a construction, or awareness of its existence, interact with geographic distribution?

REFERENCES

Adger, David. 2006. "Combinatorial variability." *Journal of Linguistics* 42:503–530.

Adger, David, and Jennifer Smith. 2005. "Variation and the minimalist program." In *Syntax and Variation: Reconciling the Biological and the Social*, ed. Leonie Cornips and Karen P Corrigan, 149–178. Philadelphia: John Benjamins Publishing Company.

Alexiadou, Artemis, Berit Gehrke, and Florian Schäfer. 2014. "The structure of adjectival participles revisited." *Lingua* 149:118–138.

Asudeh, Ash, and Ida Toivonen. 2012. "Copy raising and perception." *Natural Language and Linguistic Theory* 30:321–380.

Baker, Mark. 2001. *The atoms of language*. New York: Basic Books.

Baker, Mark C. 2008. "The macroparameter in a microparametric world." In *The Limits of Syntactic Variation*, ed. Theresa Biberauer, 351–373. Philadelphia/Amsterdam: John Benjamins.

Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther De Vogelaer, and Margreet van der Ham. 2008. *Syntactic atlas of the Dutch dialects*, volume II. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, Leonie Cornips, and Susanne van der Kleij, ed. 2002. *Syntactic microvariation*, volume II. Amsterdam: Meertens Institute Electronic Publications in Linguistics.

Barbiers, Sjef, Mario van Koppen, Hans Bennis, and Norbert Corver. 2016. "MIcrocomparative MOrphosyntactic REsearch (MIMORE): Mapping partial grammars of Flemish, Brabantish and Dutch." *Lingua* 178:5–31.

Bender, Emily M. 2001. "Syntactic variation and linguistic competence: The case of AAVE copula absence." Doctoral Dissertation, Stanford University.

Benincà, Paola, ed. 1989. *Dialect variation and the theory of grammar*. Dordrecht: Foris.

Benincà, Paola. 1994. *La variazione sintattica*. Bologna: Il Mulino.

Biggs, Alison. 2021. "The Done-State Derived Stative: A Case Study in Building Complex Eventualities in Syntax." *Syntax* 24:297–333.

Blanchette, Frances, and Chris Collins. 2019. "On the subject of negative auxiliary inversion." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 64:32–61.

Blondeau, Hélène, and Naomi Nagy. 2008. "Subordinate clause marking in Montreal Anglophone French and English." In *Social Lives in Language – Sociolinguistics and multilingual speech communities: Celebrating the work of Gillian Sankoff*, ed. Miriam Meyerhoff and Naomi Nagy, 273–313. Philadelphia: John Benjamins.

Borer, Hagit. 1984. *Parametric syntax*. Dordrecht: Foris.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R Harald Baayen. 2007. "Predicting the dative alternation." In *Cognitive foundations of interpretation*, ed. Gerlof Bouma, Irene Krämer, and Joost Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan, and Marilyn Ford. 2010. "Predicting syntax: Processing dative constructions in American and Australian varieties of English." *Language* 86:168–213.

Brook, Marisa. 2018. "Taking it up a level: Copy-raising and cascaded tiers of morphosyntactic change." *Language Variation and Change* 30:231–260.

Bruening, Benjamin. 2014. "Word formation is syntactic: Adjectival passives in English."
Natural Language and Linguistic Theory 32:363–422.

Bruening, Benjamin. 2018. "Word formation is syntactic: Raising in nominalizations." *Glossa: a journal of general linguistics* 3:102.

Bybel, Kali, and Greg Johnson. 2014. "The syntax of 'have yet to'." Paper presented at the 81st Southeastern Conference on Linguistics, March 27–29. Coastal Carolina University.

Chagnaud, Clément, Philippe Garat, Paule-Annick Davoine, Elisabetta Carpitelli, and Axel Vincent. 2017. "ShinyDialect: a cartographic tool for spatial interpolation of geolinguistic data." In *Proceedings of the 1st ACM SIGSPATIAL workshop on Geospatial Humanities*, 23–30.

Chomsky, Noam. 1995. *The Minimalist Program*. Malden, MA: MIT Press.

Chomsky, Noam. 2001. "Derivation by phase." In *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Cornips, Leonie, and Karen P Corrigan, ed. 2005. *Syntax and Variation: Reconciling the Biological and the Social*. John Benjamins.

D'Arcy, Alex. 2004. "Contextualizing St. John's youth English within the Canadian quotative system." *Journal of English Linguistics* 32:323–345.

D'Arcy, Alexandra. 2014. "Functional partitioning and possible limits on variability: A view of adjective comparison from the vernacular." *Journal of English Linguistics* 42:218–244.

de Clercq, Karen, and Guido Vanden Wyngaerd. 2019. "Negation and the functional sequence." *Natural Language & Linguistic Theory* 37:425–460.

Donati, Caterina, Francesca Foppolo, Ingrid Konrad, and Carlo Cecchetto. 2021. "Whatever His Arguments, Whatever Relatives Are Not Free Relatives: A Reply to Caponigro's Reply." *Linguistic Inquiry* early access. URL https://doi.org/10.1162/ling_a_00413.

Duncan, Daniel. 2019. "Grammars compete late: Evidence from embedded passives." *University of Pennsylvania Working Papers in Linguistics* 25:Article 11. URL https://repository.upenn.edu/pwpl/vol25/iss1/11.

Duncan, Daniel. 2021. "A note on the productivity of the Alternative Embedded Passive." *American Speech* 96:481–490.

Etxepare, Ricardo, and Ángel J. Gallego, ed. 2018–2020. *Special Issue of Glossa: Formal Approaches to Dialectal Syntax*. Glossa. URL https://www.glossa-journal.org/collections/557/.

Fruehwald, Josef, and Neil Myler. 2015. "*I'm done my homework*—Case assignment in a stative passive." *Linguistic Variation* 15:141–168.

Grieve, Jack. 2016. *Regional variation in written American English*. Cambridge: Cambridge University Press.

Harley, Heidi, and Hyun Kyoung Jung. 2015. "In Support of the P$_{have}$ Analysis of the Double Object Construction." *Linguistic Inquiry* 46:703–730. URL http://www.mitpressjournals.org/doi/10.1162/LING_a_00198.

Hartman, Jeremy. 2011. "(Non-)intervention in A-movement: Some cross-constructional and cross-linguistic considerations." *Linguistic Variation* 11:121–148.

Harves, Stephanie, and Neil Myler. 2014. "Licensing NPIs and Licensing Silence: *Have/Be Yet To* in English." *Lingua* 148:213–239.

Hasty, J. Daniel. 2014. "We might should be thinking this way: Theory and practice in the study of syntactic variation." In *Micro-syntactic variation in North American English*, ed. Raffaella Zanuttini and Laurence R. Horn, 269–293. Oxford: Oxford University Press.

Henry, Alison. 1992. "Infinitives in a *for-to* dialect." *Linguistic Inquiry* 10:279–301.

Henry, Alison. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.

Huang, Yuan, Diansheng Guo, Alice Kasakoff, and Jack Grieve. 2016. "Understanding US regional linguistic variation with Twitter data analysis." *Computers, Environment and Urban Systems* 59:244–255.

Johnson, Gregory II. 2014. "Restructuring and Infinitives: The View from Appalachia." Doctoral Dissertation, Michigan State University.

Johnson, Kyle. 1988. "Verb raising and 'have'." *McGill Working Papers in Linguistics* 156–167.

Jones, Taylor. 2015. "Toward a description of African American Vernacular English dialect regions using "Black Twitter"." *American Speech* 90:403–440.

Jones, Taylor. 2016. "Tweets as graffiti: What the reconstruction of Vulgar Latin can tell us about Black Twitter." In *English in computer-mediated communication: Variation, representation and change*, ed. Lauren Squires, 43–68. Berlin/Boston: Mouton de Gruyter.

Kaplan, Aidan. 2019. "I feel like that one complementizer is not enough." *Yale Working Papers in Grammatical Diversity* 2.

Kayne, Richard S. 1969. "The transformational cycle in French syntax." Doctoral Dissertation, Massachusetts Institute of Technology.

Kayne, Richard S. 1989. "Null Subjects and Clitic Climbing." In *The Null Subject Parameter*, ed. Osvaldo A. Jaeggli and Kenneth J. Safir, 239–261. Dordrecht: Kluwer.

Kayne, Richard S. 1991. "Romance clitics, verb movement, and PRO." *Linguistic Inquiry* 22:647–686.

Kayne, Richard S. 1997. "The English complementizer *of*." *Journal of Comparative Germanic Linguistics* 1:43–54.

Kayne, Richard S. 2005. "Some Notes on Comparative Syntax With Special Reference to English and French." In *Movement and Silence*, ed. Richard S. Kayne, 277–333. Oxford: Oxford University Press.

Kayne, Richard S. 2015. "*Once* and *twice*." *Studies in Chinese Linguistics* 36:1–20.

Kelly, Justin. 2008. "*Yet* as a negative perfect marker in English." *Snippets* 18:8–9.

Kelly, Justin. 2012. "The syntax and semantics of infinitival *yet* constructions." *Cahiers Chronos* 25:115–139.

Kendall, Tyler, Joan Bresnan, and Gerard Van Herk. 2011. "The dative alternation in African American English: Researching syntactic variation and change across sociolinguistic datasets." *Corpus Linguistics and Linguistic Theory* 7:229–244.

Kim, Chaeyoon, Sravana Reddy, James N Stanford, Ezra Wyschogrod, and Jack Grieve. 2019. "Bring on the crowd! Using online audio crowd-sourcing for large-scale new England dialectology and acoustic sociophonetics." *American Speech* 94:151–194.

Kortmann, Bernd, and Kerstin Lunkenheimer, ed. 2013. *The electronic World Atlas of Varieties of English (eWAVE)*. Max Planck Institute for Evolutionary Anthropology. URL https://ewave-atlas.org/.

Labov, William. 1966. "The effect of social mobility on linguistic behavior." *Sociological Inquiry* 36:186–203.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William, Sherry Ash, and Charles Boberg. 2006. *The Atlas of North American English*. Berlin: Mouton.

Landau, Idan. 2011. "Predication vs. aboutness in copy raising." *Natural Language & Linguistic Theory* 29:779–813.

Leinonen, Therese, Çağrı Çöltekin, and John Nerbonne. 2016. "Using Gabmap." *Lingua* 178:71–83.

Merchant, Doug. 2021. "What's in the bucket? Aspectual (non) compositionality in phrasal idioms." *Natural Language & Linguistic Theory* 40:1117–1158.

Myler, Neil. 2013. "On *coming the pub* in the North West of England: Accusative unaccusatives, dependent case and preposition incorporation." *Journal of Comparative Germanic Linguistics* 16:189–207.

Myler, Neil. 2016. "Discussion of Fruehwald and Myler (2015)." Guest lecture at Yale University, Feb. 18.

Poletto, Cecilia. 1992. "La sintassi del soggetto nei dialetti italiani settentrionali." Doctoral Dissertation, University of Padua.

Poletto, Cecilia. 1993. "Subject clitic-verb inversion in North Eastern Italian dialects." In *Syntactic theory and the dialects of Italy*, ed. Adriana Belletti, 204–251. Torino: Rosenberg & Sellier.

Poletto, Cecilia. 1996. "Three kinds of subject clitics in Basso Polesano and the theory of *pro*." In *Parameters and functional heads: Essays in comparative syntax*, ed. Adriana Belletti and Luigi Rizzi, 269–300. New York and Oxford: Oxford University Press.

Roberts, Ian. 2019. *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

Robinson, Mary, and Daniel Duncan. 2019. "Holistic approaches to syntactic variation: Wh-all questions in English." *University of Pennsylvania Working Papers in Linguistics* 25:209–218.

Röthlisberger, Melanie, and Sali A Tagliamonte. 2020. "The social embedding of a syntactic alternation: Variable particle placement in Ontario English." *Language Variation and Change* 32:317–348.

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.

Singler, John Victor. 2001. "Why you can't do a VARBRUL study of quotatives and what such a study can show us." *University of Pennsylvania Working Papers in Linguistics* 7:257–278.

Sprouse, Jon. 2007. "A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge." Doctoral dissertation, University of Maryland.

Sprouse, Jon. 2011. "A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory." *Behavior Research Methods* 43:155–167.

Srivastava, Aarohi. 2020. "Doxastic Feel Like (That)." *Yale Working Papers in Grammatical Diversity* 2:1–10.

Stoeckle, Philipp. 2016. "Horizontal and vertical variation in Swiss German morphosyntax." In *The future of dialects*, ed. Marie-Hélène Côté, Remco Knooihuizen, and John Nerbonne, 195–215. Berlin: Language Science Press.

Strelluf, Christopher. 2020. "*Needs* + PAST PARTICIPLE in regional Englishes on Twitter." *World Englishes* 39:119–134.

Strelluf, Christopher. 2022. "Regional Variation and Syntactic Derivation of Low-frequency need-passives on Twitter." *Journal of English Linguistics* 50:39–71. URL https://doi.org/10.1177/00754242211066971.

Svenonius, Peter. 2008. "Complex predicates and the functional sequence." *Nordlyd* 35:47–88.

Tagliamonte, Sali, and Alex D'Arcy. 2004. "He's like, she's like: The quotative system in Canadian youth." *Journal of Sociolinguistics* 8:493–514.

Tagliamonte, Sali, and Chris Roberts. 2005. "So weird; so cool; so innovative: The use of intensifiers in the television series Friends." *American Speech* 80:280–300.

Tagliamonte, Sali A. 2008. "So different and pretty cool! Recycling intensifiers in Toronto, Canada." *English Language & Linguistics* 12:361–394.

Tagliamonte, Sali A, and Alexandra D'Arcy. 2007. "The modals of obligation/necessity in Canadian perspective." *English World-Wide* 28:47–87.

Tagliamonte, Sali A, Alexandra D'Arcy, and Bridget Jankowski. 2010. "Social work and linguistic systems: Marking possession in Canadian English." *Language Variation and Change* 22:149–173.

Thoms, Gary, David Adger, Caroline Heycock, and Jennifer Smith. 2019. "Syntactic variation and auxiliary contraction: The surprising case of Scots." *Language* 95:421–455.

Tortora, Christina. 2018. "Evidence for Generalized Verbal Periphrasis in English." In *Boundaries Crossed, at the Interfaces of Morphosyntax, Phonology, Pragmatics and Semantics*, ed. Huba Bartos, Marcel Den Dikken, Zoltán Bánréti, and Tamás Váradi, 163–180. Springer.

Tortora, Christina, Beatrice Santorini, Frances Blanchette, and C.E.A. Diertani. 2017. *The Audio-Aligned and Parsed Corpus of Appalachian English (AAPCAppE)*. www.aapcappe.org: version 0.1.

Trousdale, Graeme, and David Adger, ed. 2007. *English Language & Linguistics: Special issue on English dialect syntax*, volume 11. Cambridge University Press.

Tyler, Matthew, and Jim Wood. 2019. "Microvariation in the *have yet to* construction." *Linguistic Variation* 19:232–279.

Van Craenenbroeck, Jeroen, Marjo van Koppen, and Antal van den Bosch. 2019. "A quantitative-theoretical analysis of syntactic microvariation: Word order in Dutch verb clusters." *Language* 95:333–370.

Van Gelderen, Elly. 2019. "Are uninterpretable features vulnerable?" *Theoretical Linguistics* 45:275–282.

Wieling, Martijn, and John Nerbonne. 2015. "Advances in dialectometry." *Annual Review of Linguistics* 1:243–264.

Willis, David. 2017. "Investigating geospatial models of the diffusion of morphosyntactic innovations: The Welsh strong second-person singular pronoun *chdi*." *Journal of Linguistic Geography* 5:41–66.

Willis, David. 2019. "Dialect syntax as a testbed for models of innovation and change: Modals and negative concord in the Syntactic Atlas of Welsh Dialects." *Glossa: a journal of general linguistics* 4:1–30.

Willis, David. 2020. "Using social-media data to investigate morphosyntactic variation and dialect syntax in a lesser-used language: Two case studies from Welsh." *Glossa: a journal of general linguistics* 5.

Wood, Jim. 2008. "*So*-inversion as Polarity Focus." *Proceedings of the 38th Western Conference on Linguistics* 19:304–317.

Wood, Jim. 2013. "Parasitic participles in the syntax of verbal *rather*." *Lingua* 137:59–87.

Wood, Jim. 2014. "Affirmative semantics with negative morphosyntax: Negative exclamatives and the New England *So AUXn't NP/DP* construction." In *Micro-Syntactic Variation in North American English*, ed. Raffaella Zanuttini and Laurence R. Horn, 71–114. Oxford: Oxford University Press.

Wood, Jim. 2019. "Quantifying geographical variation in acceptability judgments in regional American English dialect syntax." *Linguistics* 57:1367–1402.

Wood, Jim. 2022. "Microvariation in verbal *rather*." *Yale Working Papers in Grammatical Diversity* 4. Available at https://ling.auf.net/lingbuzz/006387.

Wood, Jim, Laurence R. Horn, Raffaella Zanuttini, and Luke Lindemann. 2015. "The Southern Dative Presentative meets Mechanical Turk." *American Speech* 90:291–320.

Wood, Jim, Ian Neidel, Sasha Lioutikova, Luke Lindemann, Lydia Lee, and Josephine Holubkov. 2020a. "The YGDP Mapbook: Survey Results 2015–2019." *Yale Working Papers in Grammatical Diversity* 2. URL https://elischolar.library.yale.edu/ygdp/7/.

Wood, Jim, and Raffaella Zanuttini. 2018. "Datives, Data, and Dialect Syntax in American English." *Glossa: a journal of general linguistics* 3:87.

Wood, Jim and Raffaella Zanuttini. 2023. "The syntax of English presentatives." *Language* (to appear).

Wood, Jim, Raffaella Zanuttini, Laurence R Horn, and Jason Zentz. 2020b. "Dative country: Markedness and variation in Southern dative constructions." *American Speech* 95:3–45.

Yerastov, Yuri. 2015. "A construction grammar analysis of the transitive *be* perfect in present-day Canadian English." *English Language & Linguistics* 19:157–178.

Zanuttini, Raffaella. 1991. "Syntactic properties of sentential negation. a comparative study of romance languages." Doctoral Dissertation, University of Pennsylvania, Philadelphia.

Zanuttini, Raffaella. 1995. "Dialectal variation as an insight into the structure of grammar." In *Georgetown university round table on languages and linguistics 1995*, ed. James E. Alatis, Carolyn A. Straehle, Brent Gallenberger, and Maggie Ronkin, 559–574. Washington, DC: Georgetown University Press.

Zanuttini, Raffaella. 1997. *Negation and Clausal Structure: A Comparative Study of Romance Languages*. New York: Oxford University Press.

Zanuttini, Raffaella, and Laurence R Horn, ed. 2014. *Micro-Syntactic Variation in North American English*. Oxford: Oxford University Press.

Zanuttini, Raffaella, Jim Wood, Jason Zentz, and Laurence R. Horn. 2018. "The Yale Grammatical Diversity Project: Morphosyntactic variation in North American English." *Linguistics Vanguard* 4:1–15.