needs+PAST PARTICIPLE in regional Englishes on Twitter

Christopher Strelluf

ABSTRACT

The use of *needs* with a past participle (e.g., 'The car needs washed') has been identified as a feature of the US Midland, and of Englishes in Scotland, Northern Ireland, and northern England. However, linguists have generally not been able to study *needs*+PAST in natural language data. This study reports from a large corpus of *needs*+PAST productions built from tweets in 20 US cities, 17 UK cities, and 13 other cities. It confirms *needs*+PAST as a productive feature of Scotland, Belfast, Newcastle, and the US Midland, and supports claims that the construction spread via immigration. In doing so, it validates studies based on spurious elicitations of grammaticality judgments, while also demonstrating new techniques to study low-frequency linguistic variables. It provides quantitative evidence of the extent to which a settler variety of English may leave an imprint of itself over several centuries, and of the durability of regional dialect boundaries.

KEYWORDS: past participle, *need*, twitter, regional English, low-frequency variable, grammaticality judgment

1 | INTRODUCTION

In standardized varieties of English the verb *need* can introduce passive constructions, as in (1) and (2), which are both taken from tweets:

(1)   Healthcare in America needs to be fixed. (Pittsburgh)

(2)   There's so many things that need fixing. (Pittsburgh)

In the standard construction exemplified by (1), *needs* is followed by a *be*-infinitive and past participle. In (2), which Huddleston and Pullum (2002, cited in Edelstein, 2014, p. 244) call the 'concealed passive,' *need* is followed by a present participle.

Some dialects also allow *need* to take a past participle without the intervening infinitive, as in the tweet in (3):

(3)   I suppose my car needs fixed. (Pittsburgh)

I label this construction as *needs*+PAST. *Needs*+PAST has been identified as a grammatical feature of the Midland dialect region in the United States (Ash & Boberg, 2006, pp. 293-296; Labov, Murray & Simon, 2006, p. 20; Murray, Frazer, & Simon, 1996), and of Englishes in Scotland (Brown & Millar, 1980, p. 86; Trudgill, 1983, pp. 16-17; Jamieson *et al*. 2019), Ireland (Montgomery, 1991, p. 183; Hickey 2018, map 32), and the English North (Holmes & Wilson, 2017, p. 142; Trudgill, 1983, pp. 16-17).

Needs+PAST occurs infrequently in conversation and other genres of speech, and it is difficult to elicit unconscious utterances of *needs*+PAST during sociolinguistic interviews. Studies of *needs*+PAST have therefore relied largely on surveys of judgments of grammatical

acceptability (e.g., Edelstein, 2014; Labov et al., 2006, p. 295; Murray et al., 1996), or

attestations and occurrences in dialect dictionaries, local publications, and archival materials

(e.g., Montgomery, 1991 & 1997). However, sociolinguists caution that people are unable to

assess their usage of *needs*+PAST (e.g., Labov et al., 2006, p. 296), so these judgments are

unreliable.

This article utilizes the massive body of speech-like text in Twitter to study productions

of *needs*+PAST. It examines distributions of the feature in geographic space and densities of its

occurrence, providing a large-scale study of unsolicited productions of *needs*+PAST. This study

also informs dialectological methodology. In a narrow sense, results from production data will

examine the validity of reports from surveys of grammaticality judgments, which may in turn

provide an indication of the validity of reports on other low-frequency features that have been

studied by conscious elicitation of responses. This project will also explore the usefulness of

Twitter for studying low-frequency features like *needs*+PAST, expanding the scope of

examination beyond Doyle's (2014) case study of the string *needs done* on Twitter.

2 | BACKGROUND

In American dialectology, *needs*+PAST appears to have escaped notice until Stabley's (1959)

note in the miscellany section of *American Speech* in 1959, citing frequent occurrences in the

Pennsylvania Allegheny Mountain region (p. 59). Atcheson L. Hench responded in an editor's

note that *need-made-up* was included in the 1825 *Supplement to the Etymological Dictionary of

the Scottish Language*, and suggested 'that the idiom *needs painted* had been used in Scotland

first and then had been brought to central Pennsylvania by some of the thousands of Scots who

came to the state in the eighteenth and nineteenth centuries' (Stabley, 1959, pp. 69-70).

Remarkably, these two early notes captured the tenor of observations about

needs+PAST that would follow. In line with Stabley's note, in the United States *needs*+PAST is

especially associated with the enregistered dialect of 'Pittsburghese' (e.g., Johnstone, 2013).

Popular accounts of the dialect like http://www.pittsburghese.com/ and McCool (1982) list

needs+PAST as a feature of Pittsburgh grammar, and scholars have examined acceptability

judgments of *needs*+PAST in Pittsburgh and other areas of Pennsylvania to describe the

underlying grammar of *needs*+PAST (e.g., Tenny, 1998) and to study racial divisions in local

dialects (e.g., Bloomquist, 2009).

Broader areal studies of the acceptability of *needs*+PAST have confirmed Pittsburgh as a

US locus of the feature, but have categorized it more generally as a feature of the US Midland

dialect region. Murray et al. (1996), working from a combination of survey responses, email

queries, and surreptitious recordings, report that people use (or recognize as grammatical)

needs+PAST in Ohio, Indiana, Illinois, Missouri, Iowa, Kansas, Nebraska, and South Dakota. This

band of positive responses corresponds closely to the North Midland (using McDavid's [1958]

labels) or Midland (in Labov et al., 2006; I use 'Midland' in this sense throughout this article).

Murray & Simon (2006, pp. 20-21) extend this analysis to include *needs*+PAST as one of

seventeen grammatical variables that 'define and validate a Midland variety of American

English' (p. 15). Labov et al. (2006) also surveyed judgments of the grammaticality of

needs+PAST and reached similar conclusions about the feature's areal distribution—though they

extend the *needs*+PAST isogloss west to include portions of Colorado, Wyoming, Montana, and

Idaho (p. 295). They also find islands of acceptability in Phoenix and eastern Tennessee, and exclude Iowa from the *needs*+PAST region. Maher & Wood (2011) plot a similar concentration of judgments of acceptability in the Midland, though they show at least one positive judgment in a majority of US states. As such, Stabley's note to *American Speech* took a correct step toward categorizing *needs*+PAST as a feature of English in western Pennsylvania, even if he was incorrect in confining its occurrence to that area.

Likewise, Hench's attribution of *needs*+PAST to dialects in the United Kingdom has borne out. Montgomery (1991, p. 183; 1997, p. 202) lists *needs*+PAST as a Scotch-Irish grammatical feature, and Trudgill (1983, pp. 16-17) describes it as 'normally employed in Scotland as well as in areas immediately to the south of the border,' guessing that 'perhaps four or five million native speakers of British English normally use this construction.' Holmes & Wilson (2017, p. 142) indicate that it occurs regularly in Tyneside in the English North. Murray et al. (1996, p. 258) regard it as a settled matter that *needs*+PAST was brought to the United States by Scottish and Ulster immigrants.

In spite of the confidence of these assertions of *needs*+PAST as a feature of Scottish and Ulster Englishes, *needs*+PAST has until recently received scant attention in UK dialectology. Hickey (2018, map 32) maps grammaticality judgments across Ireland for the sentence, 'My hair needs washed.' His map shows the construction to be widely accepted in Northern Ireland, and widely rejected in the Republic of Ireland, which strongly supports attribution of the feature to Ulster English. In British English, at the time of this writing, the only published attestation of *needs*+PAST is Brown & Millar's (1980, p. 86) study of auxiliary verbs in Edinburgh English, which is based on a single informant. Moreover, Brown & Millar indicate that *needs*+PAST 'is not

common in S[cottish E[nglish]' (1980, p. 86). A large new set of grammaticality judgments collected for the Scots Syntax Atlas (e.g., Jamieson *et al*. 2019) makes it clear that *needs*+PAST is widely acceptable across Scotland, and a few instances of the feature in conversations recorded for the atlas confirm that *needs*+PAST occurs in natural speech. The Scots Syntax Atlas data cast obvious doubt on Brown & Millar's claim that *needs*+PAST is uncommon in Scotland, but Brown & Millar's dismissal of *needs*+PAST as ideolectical in Scottish English at least calls into question the salience of *needs*+PAST in these varieties.

This potentially low salience of *needs*+PAST bears on the type of data linguists have been able to collect about the construction. While linguists have succeeded in mapping the areas where *needs*+PAST is judged to be acceptable and have accounted for its spread from Scottish English to Ulster English to the US Midland, this impressive body of conclusions is based on limited data. Most studies have centered on presenting *needs*+PAST constructions to informants and asking whether informants could, for example, use the construction in normal conversation. Nearly all the major studies of *needs*+PAST make use of elicited judgments, including social dialectology-oriented studies like Murray et al. (1996), Labov et al. (2006), and Bloomquist (2009), and syntax-focused studies like Tenny (1998) and Edelstein (2014).

Noteworthy exceptions that do study *needs*+PAST productions are Urley (2009), Doyle (2014), and Duncan (2019). Urley (2009) counted *needs*+PAST in newspaper articles between 1987 and 2008, based on strings of *need* followed by *repaired*, *changed*, *reminded*, *washed*, *finished*, and *taught*. Her study was limited by technology of the time, which constrained her to verbatim searches of just six strings and to the formal written genre she investigated. As a result, across tens of millions of published articles in newspapers from all 50 states over 22

years, she finds only 443 *needs*+PAST tokens. This limits the strength of her claims that

needs+PAST is spreading geographically (p. 81) and in real time (p. 68).

Duncan (2019) searched for *needs*+PAST in two online forums for Pittsburgh sports

teams. He identified 97 instances of *needs*+PAST over three days of postings and, on the basis of

needs+PAST occurring more frequently in relative clauses than other clause types (and, perhaps,

occurring less frequently in the presence of negation), argued for a revised version of the

Competing Grammars hypothesis.

Most immediately relevant to the present work, Doyle (2014) examined the

effectiveness of his tweet mapping script, SeeTweet, via a case study of the string *needs done*.

He mapped 480 instances of *needs done* in the United States, and found a high level of

agreement between his map and the acceptability *needs*+PAST judgments mapped in Labov et

al. (2006). Informants who indicated in Labov et al.'s survey that they could use *needs*+PAST

were more likely to come from areas where *needs done* occurred in tweets, and informants who

indicated they could not use the construction were less likely to come from areas where *needs*

done was tweeted (Doyle, 2014, p. 102). Doyle also looked at *needs to be done* and *needs doing*,

and found that the standard infinitival construction occurred uniformly throughout the United

States, while *needs doing* was in complementary distribution with *needs done* (pp. 104-105). His

study validated the potential to use Twitter in dialectological studies generally, and for a

needs+PAST construction specifically.

As noted above, the Scots Syntax Atlas collected a few dozen instances of *needs*+PAST

through a very large collection of recorded conversations (Jamieson *et al*. 2019). Reports of

needs+PAST have been otherwise been limited to subjective reactions because the construction

occurs infrequently in conversational speech. As Murray & Simon (2002, p. 34) note in their study of the related feature, *like*+PAST, attempting to study naturalistic productions of low-frequency grammatical features 'typically requires an inordinate amount of time and offers no guarantee of success,' as non-use 'means only that an informant has not used it yet; the construction may appear in the next sentence or […] never.' Murray et al. (1996, p. 258) acknowledge overlooking concerns about relying on conscious judgments of *needs*+PAST as 'a pragmatic decision based on the great difficulty we had in eliciting large quantities of information about [*needs*+PAST] through more traditional atlas-type methods or through relatively brief periods of free conversation.'

This reliance on the conscious judgments of informants is problematic in the case of *needs*+PAST. In mapping responses to the feature, Labov et al. (2006, p. 296) acknowledge that *needs*+PAST falls into a category of grammatical features that fall below conscious recognition. As such, informants may not notice speakers using *needs*+PAST, and may not even recognize the feature in their own grammars.

Benson (2009, p. 48; 2012, p. 232) challenges that concerns over the validity of informant judgments are 'largely anecdotal.' She defends the use of acceptability judgments for studying low-frequency features on the basis of consistency of responses within her own survey data (2009, p. 49), and points to Youmans (1986) and Bailey, Wikle, & Tillery (1997) as part of 'a growing body of evidence [that] attests to the reliability of the data and the validity of the conclusions based on acceptability judgments of morpho-syntactic features.' Chambers (1998) also shows that responses to questions in a fieldworker-administered survey are highly similar to responses to a survey conducted by mail. Importantly, though, these studies show the

reliability of responses to survey items within instruments (Benson, 2009), between instruments (Youmans, 1986), and between collection methods (Chambers. 1998). Confidence in the reliability of data across elicitation methods does not resolve concerns raised by Murray et al. (1996) and Labov et al. (2006) over the problematic nature of survey conscious judgments about a feature of which people may not be conscious.

Bailey, Wikle, & Tillery (1997, p. 57) do compare results from self-reports against observed behavior, and conclude that 'self-reports might be more valid and reliable measures of linguistic behavior than linguists have supposed.' However, their defense of self-reports for low-frequency grammatical features is based on the fact that they 'are extremely difficult to elicit' (p. 57). As such, 'any survey of these features that relies on observed behavior will surely under-report their occurrence' (pp. 57-58). In other words, Bailey, Wikle, & Tillery do not find that conscious evaluations of low-frequency grammatical features accurately reflect productions, but rather that it is difficult to observe productions of low-frequency grammatical features.

Knowledge of *needs*+PAST, then, is highly constrained by the feature's usage. It occurs infrequently in conversational speech, and therefore has been studied mostly through conscious evaluation in surveys. However, informants (perhaps because the feature occurs infrequently) may not be able to accurately report on *needs*+PAST, so responses to conscious evaluations are not necessarily reliable.

Therefore, finding ways to study productions of *needs*+PAST offers insights into knowledge of the feature specifically, but also offers insights for dialectology and dialectological methodologies more generally. Overcoming the challenges of low frequency will allow the construction to be mapped according to the production data that linguists working in Labovian

paradigms prefer over conscious evaluations. Production-based maps may be compared against

the maps drawn from grammaticality judgments to establish a baseline for the reliability of

survey responses in lieu of production data in the case of low-frequency grammatical features.

They can also be used to examine linguists' reconstructions of the historical spread of

needs+PAST, and to establish estimations of the densities of *needs*+PAST occurrences in

Englishes, which might provide insight into *needs*+PAST's unique status as a socially unmarked

and consciously unrecognized construction. The method for studying *needs*+PAST might also be

applied to other low-frequency linguistic variables.

3 | METHODS

Twitter offers a solution to the problem of low-frequency features. Because the volume of

speech-like text on Twitter is so tremendous, specific text strings can be collected on a scale

sufficient for quantitative analysis. Even when a linguistic variable occurs only as a tiny fraction

of utterances, there are so many utterances in Twitter that the tiny fraction will usually be a

large number. Moreover, because of Twitter's global reach, a researcher can collect data across

huge geographic space. Sociolinguists have demonstrated the usefulness of Twitter especially

for mapping lexical items (e.g., Austen, 2017; Doyle, 2014; Eisenstein, 2017; Eisenstein et al.,

2012; Jones, 2015; Pavalanathan & Eisenstein, 2015).

I collected tweets using the package twitteR (Gentry, 2015) in R (R Core Team, 2018).

While Doyle (2014) sampled tweets containing *needs done*, I hoped to develop methods to

study any occurrence of *needs*+PAST rather than a predetermined verbatim string. I, therefore,

sampled tweets containing any form of *need* daily from July 5 to September 4, 2018 in 20 US

cities, 17 UK cities, and 13 additional cities with large English-speaking populations. (I use the label 'global Englishes' for these 13 cities, but acknowledge the inaccuracy of that label for a set that includes Englishes of, for example, Canada, India, Nigeria, and Singapore.) In the United States, cities were selected to represent major dialect regions identified in Labov et al. (2006): the Midland, North, South, West, as well as Boston, New York City, and Philadelphia. In the United Kingdom, major population centers in England, Northern Ireland, Scotland, and Wales were sampled. I also added several smaller English cities to fill in geographic space. The 13 global English cities were selected to represent a wide geographic range of population centers where English is an official language. In the United States and United Kingdom, search radii were set to the largest single value that would not cause radii of two cities to overlap; this was 40 miles in the United States and 15 miles in the United Kingdom. Radii in global English cities were set at 25 miles.

The geocode parameter in Twitter's search API returns tweets for users who geotag tweets, but also returns tweets based on the location users enter in their profiles. User-provided locations are inexact, and it is clear that some tweets I collected for locations were not tweeted by 'native speakers' of the variety of English associated with those locations. Nevertheless, relying on user-provided location data rather than geotags was a necessary compromise, because only one or two percent of tweets are geotagged (Eisenstein, 2017; Leetaru, Wang, Cao, Padmanabhan & Shook, 2013), and *needs*+PAST occurs too infrequently to collect if sampling is limited to this tiny subset. I did not collect any other metadata on tweeters' social characteristics.

These parameters generated a corpus of 3,603,438 tweets. It was necessary to develop automated procedures to extract tweets containing *needs*+PAST, which required tagging all words in all tweets for their syntactic roles. However, Twitter text is problematic for taggers, because the brief messages provide little discursive context and because textual features of Twitter make tweets 'noisy' (e.g., Derczynski, Aswani & Bontcheva, 2013, p. 21; see also Bontcheva et al., 2013; Derczynski, Ritter, Clark & Bontcheva, 2013).

The TwitIE scripts (Derczynski et al., 2013; see also Bontcheva et al., 2013) from the GATE project improve part-of-speech tagging on tweets, because they are trained on a hand-annotated Twitter corpus and include Twitter-oriented dictionaries. However, while TwitIE performs better with tweets than do other taggers (Bontcheva et al., 2013; Derczynski et al., 2013), it is still not suited to tagging *needs*+PAST. I tested TwitIE with several dozen *needs*+PAST sentences, and nearly all participles in the construction were erroneously tagged as nouns. This is unsurprising, since TwitIE was probably not trained on a dataset that contained instances of *needs*+PAST. Even if a training set did contain instances of *needs*+PAST, they would likely occur infrequently, so the tagger would assign low likelihood that a given string was *needs*+PAST.

Marking participles as nouns is unhelpful, since it would necessitate manually removing phrases where *need* was followed by an object noun phrase. By trial and error, I found 'dummy' forms that I could substitute in the place of *need* in the corpus that would allow TwitIE to identify past participles that occurred after *need*. I then isolated all tweets where any form of *need* was immediately followed by a token tagged as a past participle.

These parameters generated an initial corpus of 7434 tweets. Of course, these solutions were not ideal. Some valid cases of *needs*+PAST were certainly excluded. For instance, sentence (4), where *either* intervenes between *needs* and the participle, was excluded from the corpus.

(4)   Your reclosable zipper bag SUCKS and needs either recalled or a real zipper on it.

        (Pittsburgh)

The parameters also returned a number of false hits:

(5)   whenever yinz need "lahd," yinz turn it allsa way up to Dad! (Pittsburgh)

(6)   We need verified impact. (Cape Town)

(7)   Woman with special needs given dream job. (Kansas City)

(8)   I just need felt or construction paper. (Philadelphia)

The tagger guessed incorrectly at the part-of-speech for the joke spelling of *loud* as <lahd> in (5). In (6), *verified* is an adjective modifying *impact*. In (7), *needs* is a noun immediately followed by a past tense verb. In (8), *need* is followed by the noun *felt*, which is mistaken for its verbal homonym. Such false hits were manually removed, reducing the corpus to 3668 tweets.

More difficult cases arose when a nominal object of *need* was elided, clefted, or otherwise displaced from a syntactic slot between *need* and a participle:

(9)   Have a jar I need opened. (Phoenix)

(10)   Selling armoire! Need gone! (Dallas)

(11)   That's the only question I need answered (Pittsburgh)

I interpret (9) with the underlying structure 'I need a jar opened,' (10) as 'I need the armoire gone,' and (11) as 'I need that question answered.' Edelstein (2014, p. 244) notes that these 'standard transitive constructions' look like *needs*+PAST, but are accepted by speakers who otherwise reject *needs*+PAST. I manually excluded standard transitive constructions, reducing the corpus to 3291 tweets.

These procedures illustrate that it is viable to cast a wide net to study productions of a low-frequency, yet highly flexible, construction like *needs*+PAST on Twitter. However, especially in early stages of research, the work will necessitate development of ad hoc procedures and close attention to quality control.

To normalize counts of *needs*+PAST tweets, I also extracted tweets where a *need* form was immediately followed by *to be* and a past participle (as in (1) above). The uniformity of this construction (at least in the United States) is attested to in Doyle (2014, pp. 104-105), so counts of *needs*+TO provide a useful baseline of 'standard' constructions in each city. The *needs*+TO corpus was checked for false hits, but proved much less error-prone than the *needs*+PAST corpus. The final *needs*+TO corpus included 44,333 tweets.

Tweets from the corpus are quoted throughout this article to exemplify *needs*+PAST constructions. All Twitter users consent to the content of their tweets being made available to anyone for any purpose, and grant Twitter the worldwide right to 'use, copy, reproduce, process, adapt, modify, publish, transmit, display, and distribute' content (Twitter Terms of

Service). While Twitter users have therefore consented to their tweets being used in research, I

have tried to maintain a layer of protection for tweeters by quoting only from accounts that are

set to be visible to the public. I do not quote tweets that begin with a call-out to a non-

commercial or non-verified account, since these might be thought of as private messages by

senders. I also avoid quoting messages that seem to refer to criminal activity or otherwise

potentially damaging information. Finally, I truncate tweets to omit text that is not relevant to

exemplifying *needs*+PAST constructions, and include only the location that returned the tweet,

as measures to lessen the degree to which tweeters' identities and other information are

revealed.

4 | RESULTS

This section presents a count-based analysis of the geographical distribution and density of

needs+PAST productions in tweets. It also examines the corpus inductively to identify additional

insights into *needs*+PAST that might be made available through Twitter corpora.

Geographical distribution

The Appendix reports the number of *needs*+PAST and *needs*+TO constructions in each city, as

well as an index, which is the proportion of *needs*+PAST counts to *needs*+TO counts (and

multiplied by 100). The indexes overwhelmingly support Montgomery's (1991 & 1997) claim of

Ulster origins for *needs*+PAST, as Belfast has the greatest index in the sample. Indeed, Belfast is

the only city where *needs*+PAST occurs more frequently than *needs*+TO. This distribution of

productions in Ireland closely reflects Hickey's (2018, map 32) mapping of grammatical

acceptability judgments. Trudgill's (1983, pp. 16-17; also Jamieson et al. 2019) claim that

needs+PAST is a feature of Scottish English and some northern English varieties (also Holmes &

Wilson, 2017, p. 142) is confirmed by high indexes in Glasgow, Edinburgh, Aberdeen, and

Newcastle.

Indexes additionally confirm Pittsburgh as the *needs*+PAST locus of the United States. Of

note, Pittsburgh's index of 38.5 in this Twitter sample is substantially higher than the 18.5

percent *needs*+PAST proportion Duncan (2019, p. 94) reported in Pittsburgh sports forums.

Other communities in the Midland dialect region also have relatively high indexes. Columbus's

index of 32.3 is nearly as high as Pittsburgh's. Indexes then drop as the Midland stretches west

to 20.1 in Indianapolis, and 11.0 in Kansas City. Cleveland, which is only 140 miles from

Columbus, but assigned to the Inland North on the basis of phonetic and phonological data in

Labov et al. (2006; see also Thomas, 2010), also shows a relatively high index of 17.5. St. Louis,

which Labov et al. (2006, pp. 276-277) identify as straddling the phonetic and phonological

patterns of the Inland North and the Midland dialect regions, appears to occupy at middle-

space with regard to *needs*+PAST, as well. St. Louis's index of 5.2 is the lowest among Midland

US cities, but greater than all other non-Midland cities besides Cleveland.

Needs+PAST has not become a part of Englishes outside the United States or the United

Kingdom. The indexes of the 13 cities included under global Englishes are very low. Moreover,

the contents of several of these *needs*+PAST tweets in global Englishes suggest tweeters are not

'native speakers' of local dialects:

(12)   People didn't need prompted to support Wales. (Hong Kong)

(13)   All voting machines need pulled immediately. (Manila)

In (12), the tweeter discusses support for British national soccer teams, and (13) is a response to news items in US media. So even the very small indexes reported for *needs*+PAST in these global Englishes likely overstate the occurrence of the construction in these speech communities. Still, some genuine productions seem to appear in global Englishes:

(14)   there are a lot of important topics since 1998 that need covered. (Toronto)

(15)   His tweets for democracy need preserved. (New Delhi)

(16)   The doctors need incentivised because […]. (Sydney)

(17)   I need put in the bin. (Dublin)

(18)   You need decried your truth-handling abilities. (Vancouver)

(19)   I need cultivated. (Singapore)

The occurrences of *needs*+PAST in Canada and Ireland are least surprising, given the presence of the construction in the United States and Northern Ireland. The feature appearing in Australia, India, and Singapore is more noteworthy. Of course, any of these tweets may be from 'non-native speakers' of local Englishes. But there seem to be enough occurrences of *needs*+PAST in Canada and Dublin, at least, to regard the construction as a very marginal feature of some local grammars, which might bear closer examination.

Indexes for the United Kingdom are mapped in Figure 1, and for roughly the eastern two-thirds of the United States in Figure 2.

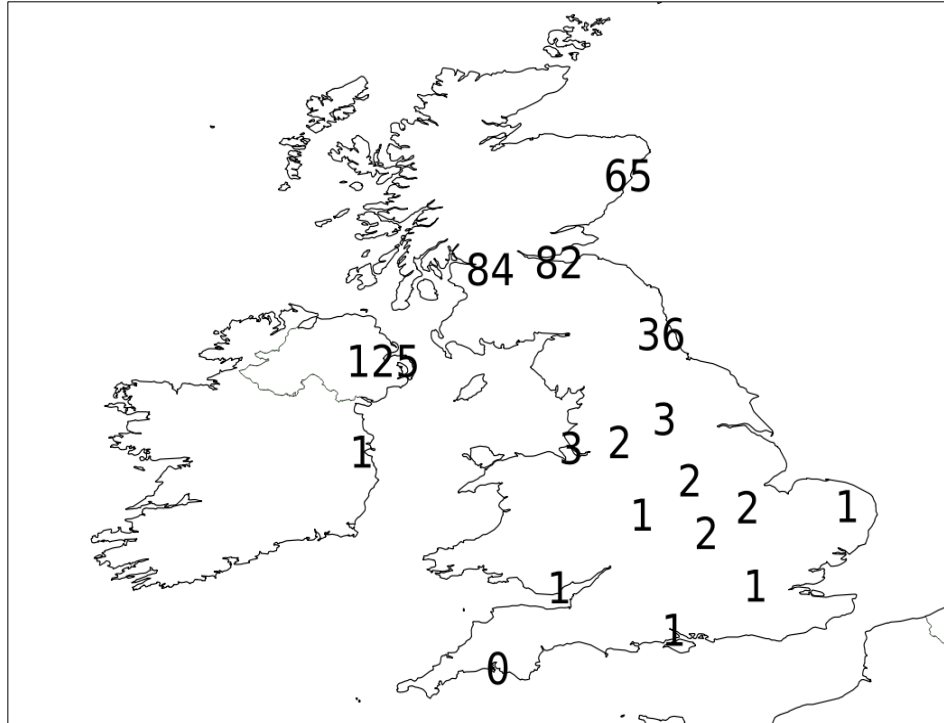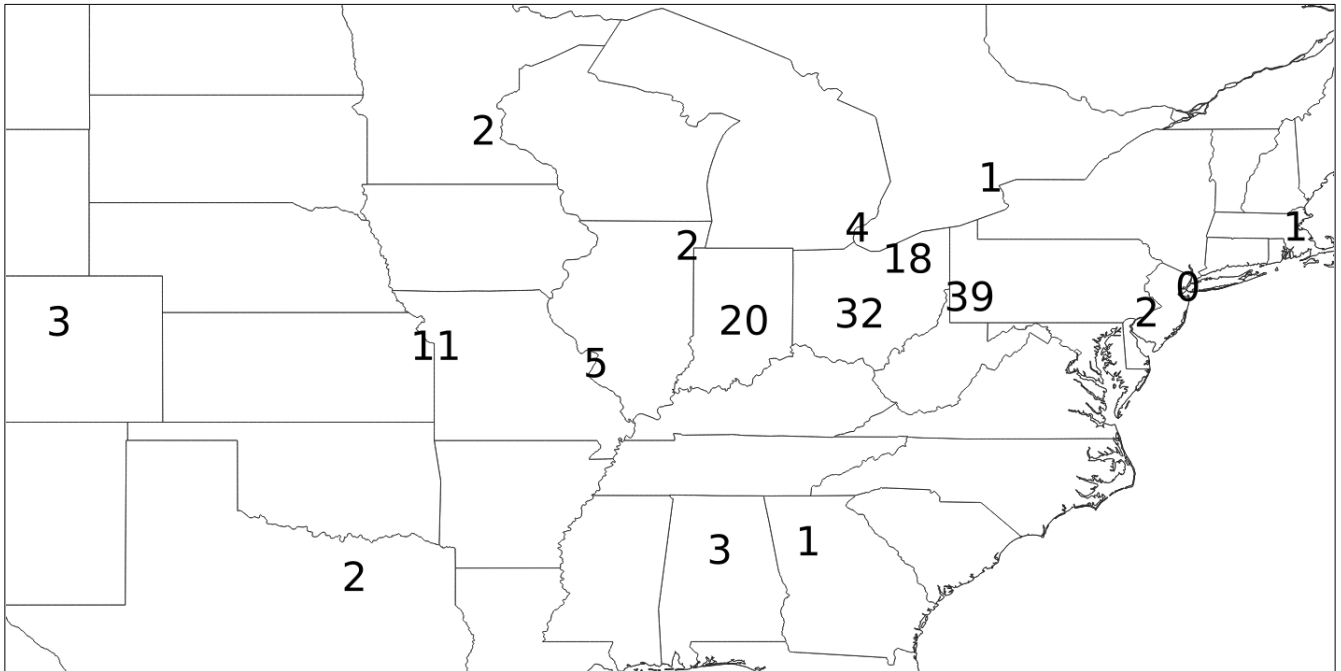**FIGURE 1** *Needs*+PAST indexes in the United Kingdom

FIGURE 2 *Needs*+PAST indexes in the United States

The geographical distribution of indexes gives a clear impression of diffusion. Proportions drop as distance from core *needs*+PAST areas increases, and quickly reach a point where *needs*+PAST constitutes a tiny portion of *needs* constructions. The pattern is apparent in Britain in Figure 1, with Glasgow and Edinburgh forming a *needs*+PAST core with indexes around 83, which decreases to 64.7 in Aberdeen 120 miles north and to 36.3 in Newcastle 100 miles south. In the United States in Figure 2, indexes diffuse westward from Pittsburgh and Columbus.

Notwithstanding the observation above that *needs*+PAST occurs occasionally in tweets from Canada and Ireland, these patterns of diffusion also reveal the remarkable durability of dialect boundaries as barriers against the spread of linguistic features. This is most obvious between Belfast and Dublin. Despite being separated by just 100 miles and sharing many commercial and social links, Belfast's index of 125.1 is opposed to Dublin's index of 1.4. Grammaticality judgments of *needs*+PAST reported by Hickey (2018, map 32) show the same dramatic division between Northern Ireland and Republic of Ireland. While (at the time of this writing), people can freely cross the border between these nations, *needs* constructions remain largely set by migrations from Scotland to Ulster during the seventeenth century (e.g., Hickey, 2007, p. 5). In Britain, similarly, English Newcastle has a much smaller index than Scottish Aberdeen, even though Newcastle is slightly closer to Edinburgh and Glasgow than Aberdeen. The English-Scottish border decreases proportions of *needs*+PAST. Moreover, Newcastle represents an obvious southern border for *needs*+PAST in Britain; indexes drop off sharply everywhere else in England and Wales.

While levels of formal political border are present between Belfast and Dublin and between Scotland and England, indexes in the United States show that historical regional dialect

boundaries may be independently sufficient to block diffusion. This is especially clear between

Pittsburgh and Philadelphia. Despite these cities both being in Pennsylvania, their *needs*+PAST

indexes are starkly differentiated, with Philadelphia having an index of just 1.8. Pittsburgh, on

the other hand, appears to be a center of gravity for *needs*+PAST diffusion in the Midland that

extends to Kansas City 850 miles to the west; *needs*+PAST does not diffuse from Pittsburgh to

Philadelphia 300 miles to the east. Pittsburgh and Philadelphia have long been recognized as

belonging to different dialect regions (e.g., McDavid, 1958), and their *needs*+PAST indexes

provide striking evidence that they remain linguistically separated. On the other hand, the

Inland North-Midland divide between Cleveland and Columbus appears to be more permeable,

and Detroit's index of 4 may also indicate some slight spillage into another nearby Inland North

community.

Generally, though, *needs*+PAST appears to adhere closely to traditional regional dialect

boundaries. Moreover, *needs*+PAST clearly seems to follow the migration of immigrants and

settlers from Scotland to Ulster Ireland to the US Midland, and then west from Pittsburgh

roughly following the National Pike. *Needs*+PAST, therefore, appears to be a lingering and

durable linguistic imprint of migrations that took place between the seventeenth and

nineteenth centuries, and one that continues to differentiate regional dialects.

The strength of this imprint might be examined through closer inspections of areas

settled by Ulster or Scottish migrants. Southern areas of New Zealand, where for many years

dialects remained /r/-ful as a result of heavy early settlement by Scottish immigrants (e.g.,

Gordon et al., 2004), are an appealing site to examine the durability of *needs*+PAST in future

work. Indeed, Maclagan & Hay (2010, p. 165) indicate that the parallel construction *wants*+PAST

occurs in New Zealand's Southland as a lingering trace of Scottish settlement. Similarly, there is great appeal to studying the transitional spaces between areas with relatively higher and lower *needs*+PAST indexes. Focused study of the area between Leeds and Glasgow/Edinburgh in the United Kingdom, between Belfast and Dublin in Ireland, and between Pittsburgh and Philadelphia in the United States might reveal exactly what the drop-off is between high- and low-frequency *needs*+PAST areas, and help draw the boundaries between dialect regions for this variable.

Additional opportunities to study social factors

The corpus does not allow for systematic investigation of social factors besides the location listed in a user's profile. However, it does present evidence of linguistic features associated with socially constructed or -meaningful categories, which may inform previous conclusions about *needs*+PAST and suggest directions for future sociolinguistic work.

Such is the case when *needs*+PAST co-occurs with features associated with African American English:

(20)   My car need cleaned already (Detroit)

(21)   It's a few asses that needs kicked. (Cleveland)

In (20), *need* shows a leveled agreement with the third-person singular subject (rather than *needs*), and (21) has *it* for the expletive subject (rather than *there*). Leveled third-person singular forms and expletive-*it* are both well-established features of African American English

(e.g., Green, 2002, pp. 36-38, pp. 80-82). These tweets then show *needs*+PAST being produced in tweets that also contain features of African American English.

Needs+PAST is not generally described as a feature of African American English. Murray et al. (1996, p. 265) cautiously suggest that *needs*+PAST is used more by European Americans than African Americans, finding that the small proportion of African Americans respondents to their grammaticality surveys rejected the construction. Murray & Simon (1999, pp. 149-150) make this claim more certainly in their research on *wants*+PAST, indicating that 'whites favor the construction significantly more than blacks,' and suggesting that *wants*+PAST and *needs*+PAST pattern similarly for social factors. On the other hand, Bloomquist (1999, pp. 38-39) finds no difference in acceptability judgments for *needs*+PAST between European and African Americans.

The co-locations of *needs*+PAST with features associated with African American English in this sample are anecdotal, but suggest that the construction might bear examination more directly as a productive feature of some varieties of African American English. It is noteworthy that many of these tweets, including examples (20) and (21), are associated with cities outside the US Midland, yielding the speculative suggestion that African American English might have helped carry *needs*+PAST beyond its Midland boundaries.

Tweets from Scotland point toward the possibility that *needs*+PAST might contribute to the construction of Scottish personae for authors. *Needs*+PAST occasionally occurs with spellings and other constructions indicative of salient features of Scottish English:

(22)   My tatties winna need watered until I get hame. (Aberdeen)

(23)   A need cuddled noo (Glasgow)

The author of (22) makes use of the Scottish *tatties* for 'potatoes,' *winna* for 'will not*,' and hame*

for 'home,' and (23) suggests Scottish English pronunciations with <a> for *I* and <noo> for *now*.

It is possible that *needs*+PAST might co-locate with these written presentations of Scottishness

as part of authors' performances of Scottish identity. This possibility could be examined by

comparing the use of eye-dialect spellings or other features indicative of Scottish English in

tweets with *needs*+PAST against *needs*+TO tweets.

As was the case with co-locations of *needs*+PAST and features of African American

English, the present study can only speculate about the potential for sociolinguistic study of co-

locations of *needs*+PAST and markers of Scottish English. Still, it is clear that using Twitter to

build a corpus of productions of *needs*+PAST and other low-frequency features creates exciting

pathways for linguistic exploration.

Conscious awareness and social evaluation of needs+PAST

Production data allow qualitative evaluation of the claim that *needs*+PAST operates below the

level of conscious awareness (e.g., Labov et al., 2006, p. 296), and is not recognized as a regional

or nonstandard feature (Murray et al., 1996, p. 266). Specifically, we can look opportunistically

for overt comments about *needs*+PAST constructions as evidence that the feature is consciously

recognized and evaluated. On the other hand, if the feature operates below the level of

consciousness, we might observe *needs*+PAST constructions in a range of communicative genres

and registers, like advertising and corporate discourse, as well as more 'vernacular' messages.

One tweet in the corpus does comment metalinguistically on *needs*+PAST:

(24)   Hey NE Ohio, has anyone pointed out that you say 'the dishes need washed,' 'the car

needs fixed,' 'the baby needs fed' instead of 'needs to be?' (Seattle)

However, the author of (24) indicates that they were unaware of *needs*+PAST until it was

pointed out to them, and subsequently noticed it occurring frequently in their own speech. The

author also links to the Yale Grammatical Diversity Project *needs*+PAST page (Maher & Wood,

2011, updated by McCoy, 2015 & Martin, 2018). The tweet, then, offers anecdotal evidence to

support the claim that *needs*+PAST operates 'below the surface.'

The sample also shows *needs*+PAST occurring in a range of professional communication

contexts:

(25)   Here's how to decide if your ducts need cleaned! (Dallas)

(26)   270 spots need filled right now. (Indianapolis)

(27)   Each form of arthritis needs treated differently. (Columbus)

These occurrences of *needs*+PAST were tweeted as organizational communiques on behalf of a

company (25), a television news station (26), and a workers' union (27). The use of *needs*+PAST

in public messages intended to represent commercial and organization interests—including in a

tweet representing broadcast media, which would be expected to have a language gatekeeping

function—attests to the construction being unstigmatized and/or unconscious. *Needs*+PAST was

also built into several recurring commercial messages, including tweets from a Columbus animal

shelter that used the template, '[dog's name] needs pulled to get out alive,' as well as solicitous

messages from an apparent sex-chat account in Edinburgh that direct-messaged phrases like 'needs sucked.' Moreover, authors occasionally quoted tweets that originally contained *needs*+TO, but rewrote them to *needs*+PAST. One such message quoted the phrase 'needs to be booted' as 'needs booted.'

The data-mining technique used to build this corpus would capture overt commentary on these messages, if it were the case that someone quoted one of these tweets and commented on it. No such metalinguistic commentary in response to *needs*+PAST, besides example (24), appears in the sample. While it is obviously problematic to draw conclusions on the basis of negative evidence, this supports sociolinguists' previous claims that *needs*+PAST passes without recognition or social stigma.

The analysis of these tweets as commentary on the availability of *needs*+PAST for conscious evaluation is necessarily unsystematic. However, it supports the longstanding claim among sociolinguists and dialectologists that *needs*+PAST operates below the level of consciousness.

Negation

Murray et al. (1996, p. 269) raise the possibility that *needs*+PAST might be less acceptable following negation than in positive scope (Duncan, 2019, p. 95, provides more empirical support for this speculation). This study present affords limited examination of this claim. The corpus includes 123 instances of *needs*+PAST in negative scope:

(28)   It does<u>n't</u> *need* reboot<u>ed</u>. (Belfast)

(29)   Wish I did<u>n't</u> *need* tak<u>en</u> care of. (Indianapolis)

These constitute just 3.7 percent of all *needs*+PAST constructions. Only 28 cities have any

instances of negated *needs*+PAST. Besides the core *needs*+PAST areas, though, several cities

where *needs*+PAST is produced infrequently do return instances of negated *needs*+PAST,

including Denver, Manchester, and Leeds. Actually, the average proportion of negated

needs+PAST outside of core areas ($\overline{x}$=14.4) is greater than in the core areas of the US Midland,

Scotland, Newcastle, and Belfast ($\overline{x}$=4.2). These numbers can be read in the context of Murray et

al.'s (1996) suggestion that *needs*+PAST might be less acceptable after negation. They confirm

that *needs*+PAST occurs much less frequently in negative scope than positive. Members of a

speech community are therefore less likely to have encountered a negated production.

However, it is not clear that there is any particular prohibition against negated *needs*+PAST.

Future research might examine negation more substantively by also examining relative

proportions of negated *needs*+TO in these communities (but see Duncan, 2019, for Pittsburgh).

5 | DISCUSSION

This paper has provided a large-scale study of *needs*+PAST based on productions in speech-like

text in Englishes spoken around the world. It has extended previous work to collect productions

data on predetermined *needs*+PAST strings (especially Doyle, 2014, but also Urley, 2009), and

identified technical approaches to sampling, tagging, and correcting a corpus of productions

that might enable research on an increasingly wide range of linguistic variables occurring in

(relatively) natural language on Twitter.

The results of this work support the claim that *needs*+PAST exists as an unconscious and unstigmatized regional variant, and confirm that *needs*+PAST is a productive feature of Englishes in Scotland, Belfast, Newcastle, and the US Midland. *Needs*+PAST productions generally follow dialectological boundaries that linguists have drawn on the basis of other types of linguistic data, such as lexical (e.g., McDavid, 1958) and phonetic and phonological (e.g., Labov et al., 2006) data in the case of the US Midland, or political and cultural barriers in the case of Scotland and Ireland. The similar patterning between *needs*+PAST and other types of linguistic variable appears to hold even in relatively nuanced cases, such as St. Louis, which seems to sit dialectologically between the US Midland and Inland North for *needs*+PAST just as it does phonetically and phonologically (e.g., Labov et al., 2006, pp. 276-278).

As such, a low-frequency—apparently socially and psychologically invisible— grammatical feature like *needs*+PAST adds a new layer of linguistic data to distinguish dialect regions from one another. *Needs*+PAST attests to the remarkable long-term stability of historical dialect boundaries, even in the face of increased contact, communication, and mobility, and despite other types of linguistic leveling (e.g., the loss of regionally distinctive lexical types). Indeed, the common occurrence of *needs*+PAST in Scotland, Northern Ireland, and the US Midland provides an enduring trace of migrations that happened hundreds of years ago as settlers moved from Scotland to Ulster Ireland, from Ulster Ireland to Pennsylvania, and from Pennsylvania to parts of the Midwest.

While this study extends knowledge of the areal distribution and density of *needs*+PAST productions, it is noteworthy that the map of *needs*+PAST and routes of diffusion are effectively the same as the ones drawn by sociolinguists and dialectologists on the basis of limited

historical attestations (e.g., Hench's response to Stabley, 1959; Montgomery, 1991, 1997) and

grammatical acceptability judgments (e.g., Labov et al., 2006, pp. 293-296; Murray et al., 1996;

Hickey, 2018; Jamieson *et al.*, 2019). Production data of *needs*+PAST on Twitter in this study

strongly confirm the conclusions of those researchers, and more broadly validate the

methodologies they employed to study *needs*+PAST. While previous work has defended the

practice of studying low-frequency linguistic features via elicitation of conscious judgments from

informants on pragmatic grounds (e.g., Murray et al., 1996; Murray & Simon, 2002) or on the

basis of reliability within instruments (e.g., Benson, 2009 & 2012) and between instruments

(e.g., Bailey, Wilke & Tillery, 1997; Chambers, 1998), the production data in this study validate

those methods as having provided a basically accurate description of *needs*+PAST in actual

language behavior. While Doyle (2014) used Labov et al. (2006) to validate the accuracy of a

sample of *needs done* collected from Twitter, I suggest that studies of productions on Twitter

like Doyle (2014) and the present article actually validate methodologies linguists have

developed as workarounds to the problem of low-frequency features. Twitter is an obviously

valuable tool for studying low-frequency linguistic variables, but we can also be reasonably

confident in continuing to rely on elicitations of conscious evaluations (see also Jamieson *et al.*,

2019).

Production data in this work also suggest some new directions for the study of

needs+PAST. For instance, despite the rigidity of historical regional dialect boundaries in

constraining *needs*+PAST, indexes in cities like Cleveland show some slippage in these borders.

Future research into transitional spaces among dialect regions might reveal important details

for understanding the construction's distribution and diffusion. Other areas of traditional

Scottish and Ulster settlement, such as southern New Zealand, also merit examination based on the strong effect of settlement patterns in the United States on *needs*+PAST. Similarly, productions of *needs*+PAST suggest that there might be value in using the traditional method of surveying acceptability judgments in areas where *needs*+PAST has not previously been recognized as a grammatical feature.

Additionally, this study has suggested that *needs*+PAST may be a feature of some African American Englishes to a greater extent than was recognized in Murray et al. (1996; also Murray & Simon, 2002), and more in line with acceptability judgments reported in Bloomquist (2009, pp. 38-39). This suggestion is based on a quantified set of observations of *needs*+PAST occurring with features associated with African American English. However, it gains incidental support from broader demographic patterns of participation in Twitter, which, in the United States, tend to skew toward over-representation of younger African Americans (e.g., Jones, 2015, pp. 405-406).

At an individual level, this research also raised the possibility that *needs*+PAST may be available to some speakers as a way to index enregistered varieties. In actuality, corpora built from Twitter offer much greater potential for sociolinguistic explorations at the level of the individual speaker. For instance, this study has reported relative proportions of *needs*+PAST in areas conceptualized as cities. However, a larger sample might allow for indexes to be calculated for individual Twitter handles, so that rates of individual speaker variation for *needs*+PAST could be examined.

Such indexes—whether at the community level as in this research or at the theoretically researchable individual level—may also help linguists understand one of the most mysterious

aspects of *needs*+PAST, which is that the construction is temporally and spatially enduring but socially and psychologically invisible. Murray et al. (1996) raise a series of questions over why accounts of *needs*+PAST conflict, why users and nonusers coexist in close proximity, how the construction has resisted sociolinguistic patterning, and how people can use the feature without being aware of it (p. 267).

Production studies may offer insights into such mysteries. Indexes of productions, for instance, might help establish the rate at which a linguistic variant must occur to persist, or the rate of occurrence at which people become aware of the variant.

The process of using computers to analyze large corpora of *needs*+PAST productions and similar variables might itself inform these questions. The computer's erroneous selection of standard transitive constructions as *needs*+PAST, as in examples (9)-(11), might reflect the experience of humans encountering or producing *needs*+PAST. Standard transitive constructions are ostensibly similar to *needs*+PAST, and it may sometimes be difficult to identify which construction is being used, as in (30), where the elided *forms* could be either the grammatical subject or object of *need signed*:

(30)   Don't forget discipline meeting forms signed. [...] Need signed and returned by

tomorrow. (St. Louis)

Standard transitive constructions occurred in every city except Georgetown and Hong Kong, and at relatively similar levels of frequency in all cities (μ=0.830, σ=0.465, using *need*+TO as a common way to normalize distributions). The surface-level similarity between the standard

transitive constructions and dialectical *needs*+PAST that trick a computer could allow

needs+PAST to slip through human mechanisms that monitor grammatical acceptability. Since

the transitive construction and *needs*+PAST both occur infrequently, there are few practical

opportunities for a person to recognize the conflicting constructions.

It is noteworthy in the context of this speculation that Kaschak & Glenberg (2004) and

Kaschak (2006) found that subjects unfamiliar with *needs*+PAST were able to understand and

generalize the construction rapidly. Perhaps this ability reflects a fact about low-frequency

constructions that are not part of a speaker's grammar but have close structural analogues in

low-frequency constructions that are part of a speaker's grammar. Additionally, these ostensibly

similar forms could interact in some way with the reduced acceptability for negated *needs*+PAST

identified by Murray et al. (1996, p. 269), if an interceding negative particle somehow blocked

the standard transitive workaround.

These possibilities require psycholinguistically informed experiments to test. They would

have interesting potential analogues in other frequent constructions like BE+PAST for

progressive aspect in Northern British Englishes ('I was sat at my desk'; e.g., Stange, 2016), or

want+PREPOSITION ('The baby wants up'; e.g., Murray & Simon, 2006, pp. 26-27; Benson,

2009) and positive-*anymore* ('There's a lot to do downtown anymore'; e.g., Youmans, 1986;

Strelluf, 2019) in the US Midland.

Whether such speculative suggestions bear out, it is clear that social media offer exciting

possibilities to study productions of low-frequency linguistic variables, which have previously

proven difficult to study. This article demonstrates some technical approaches that may be

employed for such work. These approaches offer new directions for the study of low-frequency

linguistic features in varieties of Englishes around the world, and for our knowledge of Englishes

and of language more broadly.

REFERENCES

Austen, M. (2017). 'Put the groceries up': Comparing black and white regional variation.

   American Speech, 92, 298-320. doi:10.1215/00031283-4312064

Bailey, G., Wilke, T., & Tillery, J. (1997). The effects of methods on results in dialectology. *English*

   World-Wide, 18, 35-63. doi:10.1075/eww.18.1.03bai

Benson, E. J. (2009). Everyone wants *in*: *Want*+prepositional adverb in the Midland and beyond.

   Journal of English Linguistics, 37, 28-60.

Benson, E. J. (2012). *Need*+prepositional adverb in the Midland: Another feature *needs in*.

   Journal of English Linguistics, 40, 224-255.

Bloomquist, J. (2009). Dialect differences in central Pennsylvania: Regional dialect use and

   adaptation by African Americans in the lower Susquehanna Valley. *American Speech, 84*,

   27-47.

Bontcheva, K., Derczynski, L., Funk, A., Greenwood, M. A., Maynard, D., & Aswani, N. (2013).

   TwitIE: An open-source information extraction pipeline for microblog text. In G.

   Angelova, K. Bontcheva, & R. Mitkov (Eds.), *Proceedings of the International Conference*

   on Recent Advances in Natural Language Processing (RANLP 2013) (pp. 83-90). Hissar,

   Bulgaria: RANLP. Retrieved from https://gate.ac.uk/sale/ranlp2013/twitie/twitie-

   ranlp2013.pdf

Brown, K., & Millar, M. (1980). Auxiliary verbs in Edinburgh speech. *Transactions of the*

   Philological Society, 78, 81-133.

Chambers, J.K. (1998). Inferring data from a postal questionnaire. *Journal of English Linguistics,*

   26, 222-246.

Derczynski, L., Maynard, D., Aswani, N., & Bontcheva, K. (2013a). Microblog-genre noise and

impact on semantic annotation accuracy. In G. Stumme, & A. Hotho (Eds.), *Proceedings*

of the 24th ACM Conference on Hypertext and Social Media (HT '13) (pp. 21-30). New

York: Association for Computing Machinery.

Derczynski, L., Ritter, A., Clark, S., & Bontcheva, K. (2013b). Twitter part-of-speech tagging for all:

Overcoming sparse and noisy data. In G. Angelova, K. Bontcheva, & R. Mitkov (Eds.),

Proceedings of the International Conference on Recent Advances in Natural Language

Processing (RANLP 2013) (pp. 198-206). Hissar, Bulgaria: RANLP. Retrieved from

http://www.derczynski.com/sheffield/papers/twitter_pos.pdf

Doyle, G. (2014). Mapping dialectal variation by querying social media. In S. Wintner, S.

Goldwater, & S. Riezler (eds.), *Proceedings of the 14th Conference of the European*

Chapter of the Association for Computational Linguistics (pp. 98-106). Gothenburg,

Sweden: Association for Computational Linguistics. Retrieved from

http://aclweb.org/anthology/E/E14/E14-1011.pdf

Duncan, D. (2019). Grammars compete late: Evidence from embedded passives. *University of*

Pennsylvania Working Papers in Linguistics, 25, 89-98. Retrieved from

https://repository.upenn.edu/pwpl/vol25/iss1/11/.

Edelstein, E. (2014). This syntax needs studied. In R. Zanuttini, & L. Horn (Eds.), *Micro-syntactic*

variation in North American English (pp. 242-268). Oxford: Oxford University Press.

Eisenstein, J. (2017). Written dialect variation in online social media. In C. Boberg, J. Nerbonne,

& D. Watt (Eds.), *Handbook of dialectology* (pp. 368-383). Hoboken, NJ: Wiley.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2012). Diffusion of lexical change in social

      media. *PLOS ONE, 9*, e11 3114. doi:10.1371/journal.pone.0113114

Gentry, J. (2015). twitteR: R Based Twitter Client. R package version 1.1.9. Retrieved from

      https://CRAN.R-project.org/package=twitteR.

Gordon, E., Campbell, L., Hay, J., Maclagan M., Sudbury, A., & Trudgill, P. (2004). *New Zealand*

      English: Its origins and evolution. Cambridge: Cambridge University Press.

Green, L. J. (2002). *African American English: A linguistic introduction*. Cambridge: Cambridge

      University Press.

Hickey, R. (2007). *Irish English: History and present-day forms*. Cambridge: Cambridge University

      Press.

Hickey, R. (2018). A Survey of Irish English Usage – Choropleth maps. Retrieved from

      https://www.uni-due.de/IERC/

Holmes, J., & Wilson, N. (2017). *An introduction to sociolinguistics* (5th ed.). London: Routledge.

Huddleston, R. & Pullum, G. K. (2002). *The Cambridge grammar of the English language*.

      Cambridge: Cambridge University Press.

Jamieson, E, Chien, S., Thoms, G., Adger, D., Heycock, C., & Smith, J. (2019, September). *When*

      intuitions (don't) fail: Sociosyntax in the analysis of Scots. Paper presented at UKLVC,

      London.

Johnstone, B. (2013). *Speaking Pittsburghese: The story of a dialect*. Oxford: Oxford University

      Press.

Jones, T. (2015). Toward a description of African American Vernacular English dialect regions

      using 'Black Twitter.' *American Speech, 90*, 403-440. doi:10.1215/00031283-3442117

Kaschak, M. P. (2006). What this construction needs is generalized. *Memory & Cognition, 34*, 368-379.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133*, 450-467.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: De Gruyter.

Leetaru, K. H., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday, 18*. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/4366/3654.

Maclagan, M., & Hay, J. (2010). Sociolinguistics in New Zealand. In M. J. Bell (Ed.), *The Routledge handbook of sociolinguistics around the world* (pp. 159-169). London: Routledge.

Maher, Z., & Wood, J. (2011). Needs washed. *Yale Grammatical Diversity Project: English in North America*. Updated by T. McCoy (2015) and K. Martin (2018). Retrieved from http://ygdp.yale.edu/phenomena/needs-washed.

McCool, S. (1982). *Sam McCool's new Pittsburghese: How to speak like a Pittsburgher*. Pittsburgh: Hayford Press.

McDavid, R. I. (1958). The dialects of American English. In W. N. Francis (Ed.), *The structure of American English* (pp. 480-543). New York: Ronald Press.

Montgomery, M. (1991). The roots of Appalachian English: Scotch-Irish or British Southern? *Journal of the Appalachian Studies Association, 3*, 177-191.

Montgomery, M. B. (1997). The Scotch-Irish element in Appalachian English: How broad? How deep? In H. T. Blethen & C. W. Wood, Jr. (Eds.), *Ulster and North America: Transatlantic perspectives on the Scotch-Irish* (pp. 189-212). Tuscaloosa: University of Alabama Press.

Murray, T. E., Frazer, T.C., & Simon, B. L. (1996). *Need*+past participle in American English. *American Speech, 71*, 255-271.

Murray, T. E., & Simon, B. L. (1999). *Want*+past participle in American English. *American Speech, 74*, 140-164.

Murray, T. E., & Simon, B. L. (2002). At the intersection of regional and social dialects: The case of *like*+past participle in American English. *American Speech, 77*, 32-68.

Murray, T. E., & Simon, B. L. (2006). What is dialect?: Revisiting the Midland. In T. E. Murray & B. L. Simon (Eds.), *Language variation and change in the American Midland: A new look at "Heartland" English* (pp. 1-30). Amsterdam: John Benjamins.

Pavalanathan, U., & Eisenstein, J. (2015). Audience-modulated variation in online social media. *American Speech, 90*, 187-213. doi:10.1215/00031283-3130324

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Stabley, R. R. (1959). 'Needs painted,' etc., in Western Pennsylvania. *American Speech, 34*, 69-70.

Stange, U. (2016). *I was sat there talking all night*: A corpus-based study on factors governing intra-dialectal variation in British English. *English Language and Linguistics, 20*, 511–531. doi:10.1017/S1360674316000319.

Strelluf, C. (2019). *Anymore*, it's on Twitter: Positive *anymore*, American regional dialects, and

   polarity licensing in tweets. *American Speech, 94*, 313-351. doi: 10.1215/00031283-

   7587883.

Tenny, C. (1998). Psych verbs and verbal passives in Pittsburghese. *Linguistics, 36*, 591-597.

Thomas, E. R. (2010). A longitudinal analysis of the durability of the Northern-Midland dialect

   boundary in Ohio. *American Speech, 85*, 375-430.

Trudgill, P. (1983). *On dialect: Social and geographical perspectives*. Oxford: Blackwell.

Twitter Terms of Service. (2018). Twitter. Retrieved from https://twitter.com/en/tos.

Urley, K. S. (2009). Dinner *needs* cooked, groceries *need* bought, diapers *need* changed, kids

   need bathed: Tracking the progress of *need*+past participle across the United States.

   Muncie, IN: Ball State University Master of Arts Thesis.

Youmans, G. (1986). Any more on *anymore*: Evidence from a Missouri regional dialect survey.

   American Speech, 61, 61-75.

APPENDIX

United Kingdom cities

City	needs+TO	needs+PAST	Index
Aberdeen, Scotland	116	75	64.7
Belfast, NI	175	219	125.1
Birmingham, England	674	6	0.9
Cardiff, Wales	456	6	1.3
Edinburgh, Scotland	510	420	82.4
Glasgow, Scotland	826	691	83.7
Leeds, England	719	22	3.1
Liverpool, England	957	26	2.7
London, England	1501	22	1.5

Manchester, England	1580	35	2.2
Newcastle, England	570	207	36.3
Northampton, England	126	2	1.6
Norwich, England	227	2	0.9
Nottingham, England	644	14	2.2
Peterborough, England	97	2	2.1
Plymouth, England	102	0	0.0
Southampton, England	272	2	0.7

United States cities

City	needs+TO	needs+PAST	Index
Atlanta, Georgia	854	12	1.4
Birmingham, Alabama	585	17	2.9
Boston, Massachusetts	1940	25	1.3
Chicago, Illinois	1531	23	1.5
Cleveland, Ohio	644	113	17.5
Columbus, Ohio	603	195	32.3
Dallas, Texas	1608	28	1.7
Denver, Colorado	754	21	2.8
Detroit, Michigan	732	29	4.0
Indianapolis, Indiana	1006	202	20.1
Kansas City, Missouri	915	101	11.0
Los Angeles, California	533	11	2.1
Minneapolis, Minnesota	1119	17	1.5
New York City, New York	1045	4	0.4
Philadelphia, Pennsylvania	1917	35	1.8
Phoenix, Arizona	2008	54	2.7
Pittsburgh, Pennsylvania	1194	460	38.5
San Francisco, California	1677	8	0.5
Seattle, Washington	2318	59	2.5
St Louis, Missouri	1190	62	5.2

Global Englishes cities

City	needs+TO	needs+PAST	Index
Auckland, New Zealand	276	1	0.4
Cape Town, South Africa	499	0	0.0
Dublin, Ireland	1111	15	1.4
Georgetown, Guyana	4	0	0.0
Hong Kong, China	184	1	0.5
Islamabad, Pakistan	402	0	0.0
Lagos, Nigeria	620	2	0.3
Manila, Philippines	528	4	0.8
New Delhi, India	1702	2	0.1
Singapore	543	1	0.2
Sydney, Australia	1564	6	0.4
Toronto, Canada	1966	19	1.0
Vancouver, Canada	1209	13	1.1