Regional Variation and Syntactic Derivation of Low-frequency NEED-passives on Twitter

Christopher Strelluf¹

Abstract
This paper examines constructions formed by the verb NEED taking a passivized complement. While previous dialectological, sociolinguistic, and micro-syntactic analyses have focused primarily on the past-participle complement (NEED+ED) as a regional syntactic variable, this study expands the purview of NEED-passives to examine gerund-participle (NEED+ING) and infinitival (NEED+TO) complements. It also looks beyond purported NEED+ED regions to examine NEED-passive variation in Englishes spoken around the world. Data from Twitter confirm previous findings that NEED+ED is a productive feature of the US Midland, Scotland, Northern Ireland, and Tyneside, England. However, tweets also show that NEED+ING is produced disproportionately frequently in England and Wales. These results reveal a more complex pattern of NEED-passive variation in global Englishes than has previously been reported. Additionally, a transitive construction formed with NEED as a matrix verb is examined and found to co-vary regionally with NEED+ING. Syntactic analyses of tweets reveal similarities in the ways that NEED+ED and NEED+ING vary with NEED+TO. These findings lead to a proposed syntactic analysis that NEED+ED and NEED+ING share the same derivational structure. More generally, the work argues for greater attention in linguistic research to low-frequency features.

¹University of Warwick, Coventry, UK

Corresponding Author:
Christopher Strelluf, Applied Linguistics, University of Warwick, S1.74 Social Sciences Building, Coventry CV4 7AL, UK.
Email: c.strelluf@warwick.ac.uk
Keywords
alternative embedded passive, concealed passive, participle, gerund, catenative, micro-syntax, Midland, Scotland, England, Britain

1. Introduction

Varieties of English may allow the verb NEED to take a passive complement in any of several constructions. Options are illustrated in (1)-(6), which are taken from tweets by handles associated with the parenthesized cities. In each example, the NEED lexeme is italicized and the passive complement is underlined.¹

(1) That rule *needs* to be changed. (San Francisco)
(2) Basketball is ridiculous and *needs* fixing. (Boston)
(3) Kauffman Stadium *needs* moved up just for having fountains. (Kansas City)
(4) There’s one question we *need* to be answered. (Seattle)
(5) That’s what you *need* doing as well. (Birmingham, UK)
(6) I have some simple jobs I *need* done. (Philadelphia)

The infinitival be and past participle in (1) follows a regular pattern for English passivization. In this paper, I refer to this construction as NEED+TO. In (2), which I label NEED+ING, NEED takes a gerund-participle complement, rendering the same meaning as an equivalent NEED+TO construction.² In (3), NEED takes a past participle complement, which also has the same meaning as the equivalent NEED+TO sentence. Examples (4)-(6) are transitive constructions, where NEED takes an active subject and a non-finite clause as a direct object. In these examples, the passive object of the non-finite clause has been dislocated to the left (i.e., “some simple jobs I *need* done” has the same underlying syntax as “I *need* some simple jobs done”). I label these transitive-NEED+TO (4), transitive-NEED+ING (5), and transitive-NEED+ED (6).

Previous research on NEED-passives has focused primarily on NEED+ED as a regionally restricted variant (e.g., Murray, Frazer & Simon 1996; Tenny 1998; Edelstein 2014). The present study expands the purview of NEED-passive variation research to include NEED+ING and NEED+TO as variants in their own right and explores NEED-passive variation across a range of global Englishes. This examination reveals that NEED-passive variation is not limited to a difference between regional grammars that allow NEED+ED and those that do not. Rather, NEED+ING displays regional variability that mirrors the variability of NEED+ED and, in many Englishes, both NEED+ED and NEED+ING are only marginally productive. This project further explores a small, opportunistically collected corpus of transitive-NEED, which reveals regional variability in these constructions too. As such, this study recasts NEED-passives as a system of syntactic variability in Englishes worldwide.

The syntax of NEED-passives is further examined by reviewing derivational accounts for each NEED-passive, focusing especially on Edelstein’s (2014) analysis that NEED+ED is formed from a distinctive syntactic operation. I apply syntactic tests that evinced...
NEED+ED’s unique derivational structure in Edelstein’s (2014) study, and find that NEED+ING follows the same syntactic constraints as NEED+ED. This provides evidence for a new derivational account of NEED+ING as being generated by the same syntax as NEED+ED. I suggest that this common derivational structure between NEED+ED and NEED+ING may explain several puzzling features of NEED-passives (see section 2).

Put together, these findings highlight the value of increased and expanded examination of low-frequency variables like NEED-passives in dialectology, sociolinguistics, micro-syntax, and other areas of linguistic research. I find that the NEED-passive system of English has been mis-analyzed as a consequence of the challenges that low-frequency features pose for linguists engaged in empirical study of productions of natural language. In this study, the tremendous volume of naturalistic language on Twitter makes it possible to extensively examine a variable that occurs too infrequently in traditional natural-language corpora to allow quantitative analysis. I demonstrate the value of social media as a resource to afford increased attention to low-frequency variables, and argue that linguistic theory and knowledge of language will be enriched through such work.

2. Background

This section provides context for the dialectological, sociolinguistic, and micro-syntactic exploration of NEED-passives. I describe previous research on the distribution and syntax of NEED+ED, NEED+ING, and transitive-NEED. I do not provide focused discussion of NEED+TO because, in the context of NEED-passive variation, NEED+TO has primarily been positioned as a standard alternative to NEED+ED. The section concludes with discussion of low-frequency features and the exigency of examining them.

2.1. NEED+ED and Other Alternative Embedded Passives

NEED+ED has been described as a syntactic variant of the US Midland (Murray, Frazer & Simon 1996; Murray & Simon 2006; Labov, Ash & Boberg 2006:294-295; Maher & Wood 2011), Scotland (Jamieson et al. 2019; Smith et al. 2019), Northern Ireland (Hickey 2018), and Tyneside, England (Trudgill 1983:16-17; Holmes & Wilson 2017:142). Maclagan and Hay (2010:165) also indicated that the construction is present in areas of New Zealand that were settled primarily by migrants from Scotland. Strelluf (2020) mapped the occurrence of NEED+ED in tweets from fifty cities in the United States, United Kingdom, and elsewhere in the world. The production data confirmed the dialectological mapping of NEED+ED that previous studies had created from surveys of elicited judgments about grammatical acceptability, while also showing regional and intra-regional differences in the frequency of NEED+ED relative to NEED+TO. In particular, Strelluf (2020) showed that NEED+ED was a more robust feature of Englishes in Northern Ireland and Scotland than in the US Midland where the feature had been most often previously researched, and that in the US Midland, usage was concentrated in Pittsburgh and then dissipated as the Midland extended west.
Edelstein (2014) used the label “alternative embedded passive” (AEP) for NEED+ED and a small set of verbs that may take past participles as passive complements. It is well established that three verbs license the AEP: NEED (Murray, Frazer & Simon 1996), WANT (e.g., “The cat wants fed”; Murray & Simon 1999), and LIKE (e.g., “The cat likes cuddled”; Murray & Simon 2002). Murray and Simon (2002:59) identified an implicational scale among AEP NEED, WANT, and LIKE, with LIKE+ED being acceptable only to speakers who also accept WANT+ED, and WANT+ED in turn only being acceptable to speakers who accept NEED+ED. Edelstein (2014:258-259) confirmed this implicational scale in acceptability judgments in Pittsburgh. Additional verbs that take past participle passive complements include could use (Tenny 1998:596), could stand (Doyle 2012), DESERVE, LOVE, HATE, WISH, and HOPE (Duncan 2021).

In their examination of WANT, Murray and Simon (1999:157) noted the possibility that WANT+ED (and by extension to the broader class of AEP verbs, NEED+ED) may be merely an elided form of WANT+TO. However, they argued that “too many of our respondents use [WANT+ED] exclusively and unconsciously […] and, in fact, these speakers often object to [WANT+TO] not just as a matter of register but as a matter of grammar,” and concluded that WANT+TO and WANT+ED were syntactically distinct constructions. Tenny (1998:596) likewise argued that NEED+ED is derived by a different syntax from NEED+TO. Tenny (1998:596) identified differing constraints on NEED+TO and NEED+ED among speakers in Pittsburgh, including that NEED+ED takes a much more limited set of verbs as passive complements than NEED+TO.

Edelstein (2014) agreed that the AEP and “standard embedded passive” (i.e., NEED+TO) are syntactically distinct forms. Her analysis hinged on the distinction between raising constructions and object control predicates in the argument structure of matrix verbs and non-finite clauses. The verbs WANT and LIKE are normally control predicates, where the matrix verb assigns a thematic role to the grammatical subject of the matrix clause. However, Edelstein (2014) found evidence in acceptability judgments that when WANT and LIKE appear as the matrix verb in the AEP, they are raising constructions, where the verb in the passive complement assigns the thematic role to the grammatical subject in the matrix clause. The change in argument structure for matrix verbs when they appear in the AEP versus when they appear in the standard embedded passive provides compelling evidence that the AEP is derivationally distinct from the standard embedded passive.

Furthermore, Edelstein (2014:258-259) noted that the rates at which respondents judged NEED+ED, WANT+ED, and LIKE+ED to be acceptable generally mirrored the extent to which these verbs are normally raising constructions or object control predicates: NEED is normally raising; WANT is normally a control predicate, but sometimes allows a raising reading; LIKE is almost always a control predicate. As such, Edelstein’s (2014) account of the AEP as a raising construction offers a syntactic explanation for the NEED > WANT > LIKE implicational scale identified by Murray and Simon (2002): NEED is inherently raising, so NEED+ED naturally fits into the raising AEP; WANT is potentially raising, so some speakers can use WANT+ED in the raising AEP; LIKE is rarely raising, so just a few speakers can make LIKE+ED work in the raising AEP.
The derivation Edelstein (2014:265) proposed for the AEP is reproduced for 
NEED+ED in Figure 1. It is juxtaposed against a derivation of NEED+TO in Figure 2, 
which follows the syntax described for raising constructions in Adger (2003:318, and 
personal communication, 19-20 February 2021). In both diagrams, angle brackets 
indicate where lexical items initially merge into the derivation before moving to the 
position where they are pronounced. I have simplified some details of Adger’s (2003) 
derivation. Figures were created with LingTree by SIL International (2020).

Edelstein’s (2014) syntactic analysis depicted in Figure 1 proposes that the matrix 
verb in AEP constructions directly selects an aspect phrase (AspP) that assigns passive 
morphology to the verb in the passive complement. This contrasts with the standard 
eMBEDDED PASSIVE SYNTAX FOR A RAISING CONSTRUCTION IN FIGURE 2, WHERE THE MATRIX VERB 
SELECTS A TENSE PHRASE (TP) AS A COMPLEMENT, WHICH IN TURN SELECTS A PASSIVE PHRASE THAT 
ASSIGNS PASSIVE MORPHOLOGY.

Edelstein (2014:265) noted that her account “puts the matrix verb [of the AEP] in a 
more local relationship with the participle than occurs when additional structure is 
present,” as in the standard embedded passive. This local relationship accounts for 
several syntactic characteristics of the AEP. In particular, both Tenny (1998) and 
Edelstein (2014) analyzed that the participle complement to the AEP is always verbal, 
while the complement to standard embedded passives may be either verbal or ad-
jectival. Edelstein (2014:265) argued that “it follows that the AspP, which determines 
this categorization, should be directly selected by the matrix verb, with no intervening 
structure.” The derivation in Figure 1 also explains why respondents to Edelstein’s 
(2014) survey of grammatical acceptability judgments rejected AEP constructions 
where negation intervened between a matrix verb and passive complement; unlike the 
TP in Figure 2, the AspP in the AEP does not allow a projection for negation.

2.2. NEED+ING and Other Concealed Passives

Huddleston and Pullum (2002:1999-1200) included NEED+ING under the heading of 
the “concealed passive.” Concealed passives are formed by a matrix verb taking as a 
complement a non-finite clause that contains a gerund-participle verb. Figure 3 
suggests a derivational syntax for NEED+ING following Huddleston and Pullum’s 
(2002) description (and treating NEED+ING as a raising construction like NEED+TO, so 
that its non-finite phrasal complement will be a tense phrase).

While Edelstein (2014:244) explicitly differentiated the concealed passive from the 
AEP, there are striking parallels between the two constructions. In particular, 
Huddleston and Pullum (2002:1200, 1231) categorized NEED, WANT, DESERVE, and 
REQUIRE as being able to take the concealed passive. Edelstein (2014:244) cited DESERVE 
and REQUIRE as evidence that the concealed passive “allows a wider array of matrix 
verbs” than the AEP. However, as the list of attested AEP matrix verbs above shows, 
“the AEP is more productive than the literature describes” (Duncan 2019:3; see also 
Duncan 2021). Generally, it seems that the AEP and concealed passive can be formed 
from an identical set of matrix verbs.4
Figure 1. Derivational Structure of NEED+ED Recreated from Edelstein (2014:265) for “The cat needs fed”

Figure 2. Derivational Structure of NEED+TO Recreated from Adger (2003:318) for “The cat needs to be fed”
While the full extent of overlap between the matrix verbs that allow the concealed passive and AEP has not, to my knowledge, been explicitly acknowledged in previous literature, several studies have dealt with the AEP and concealed passive as being in a relationship of complementary distribution. Murray, Frazer, and Simon (1996:266), for example, implied this as they described NEED+ED as invisible to sociolinguistic evaluation for speakers who use it, indicating that when NEED+ED speakers are presented with NEED+ING as a grammatical alternative, “they reject it as ‘ungrammatical,’ ‘funny,’ or ‘odd,’ just as they reject [NEED+TO] as ‘too formal.’” Murray & Simon (1999:158) also indicated that their classroom surveys of WANT+ED qualitatively showed it to be in complementary distribution with WANT+ING. Complementary distribution is similarly implied in Labov, Ash, and Boberg’s (2006:293) description of NEED+ED as an option “where other dialects use [NEED+ING] or [NEED+TO].”

Following the implication of complementarity in Murray, Frazer, and Simon (1996), Doyle (2014:104) searched for tweets containing the strings needs to be done, needs done, and needs doing. He mapped locations of tweets from the continental United States that contained these three strings and found that the NEED+TO string “is acceptable in most locations,” whereas [NEED+ING] “is strongest in the areas where [NEED+ED] is not used,” providing empirical evidence that NEED+ING and NEED+ED are complementary variants in the United States (Doyle 2014:104-105).

Doyle’s (2014) study is unique in conceptualizing NEED+ING as a variable analogous to NEED+ED. Other researchers generally seem to have taken it for granted that NEED+ED was the variant of interest, while NEED+ING was unexceptional. This approach is revealed not only in lack of syntactic examinations of NEED+ING equivalent

![Figure 3. Derivational Structure of NEED+ING based on Description in Huddleston and Pullum (2002:1999-1200) for “The cat needs feeding”](image)
to those of \textsc{need+ed} (e.g., Tenny 1998; Edelstein 2014), but also in small rhetorical moves, such as Murray, Frazer, and Simon’s (1996:266) positioning of \textsc{need+ed} as a “regional” alternative to \textsc{need+ing}, and Labov, Ash, and Boberg’s (2006:293) juxtaposition of \textsc{need+ed} against “other dialects” that use \textsc{need+ing} or \textsc{need+to}. Descriptions of the concealed passive in Huddleston and Pullum (2002:1199-1200, 1231) give no indication that the construction is anything but standard across Englishes. De Smet (2013:85; see also 2014:232) noted variation in written Englishes between \textsc{want+to} and \textsc{want+ing} as early as the fourteenth century in a broader examination of the collapse of the gerund/participle distinction in English, which naturally positions \textsc{want+ing} as indicative of larger patterns in English.

2.3. Transitive-\textsc{need}

De Smet (2014:85-86) found the earliest corpus attestations of both transitive-\textsc{need+ed} and transitive-\textsc{need+ing} in the beginning of the twentieth century. Despite the recent appearance of the constructions in English, both are sufficiently established in Englishes to be noted in grammars like Quirk, Greenbaum, Leech, and Svartvik (1985:1207) and Huddleston and Pullum (2002:1206, 1245). Quirk, Greenbaum, Leech, and Svartvik (1985:1207) described transitive-\textsc{need+ed} as “a raised object followed by an -ed participle clause,” exemplified by the sentence reprinted here as (7). They gave no indication that the construction is anything but standardly available across Englishes.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{A figure related to the text content.}
\end{figure}

\begin{equation}
I \text{ want/need} \text{ this watch repaired immediately. (Example [16.54ii] from Quirk, Greenbaum, Leech & Svartvik 1985:1207)}
\end{equation}

Huddleston and Pullum (2002:1245) cited both transitive-\textsc{need+ed} and transitive-\textsc{need+ing} as the “concealed passive in a complex catenative construction” with the examples reprinted as (8) and (9).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{A figure related to the text content.}
\end{figure}

\begin{equation}
\text{He needs/wants his hair cut. (Extracted from example [60iv] in Huddleston & Pullum 2002:1245)}
\end{equation}

\begin{equation}
\text{He needs/wants his hair cutting. (Extracted from example [60iv] in Huddleston & Pullum 2002:1245)}
\end{equation}

However, in contrast to Quirk, Greenbaum, Leech, and Svartvik (1985), Huddleston and Pullum (2002:1245) described transitive-\textsc{need+ed} in (8) “as an alternant of the concealed passive,” which is “restricted to certain regional dialects such as Scottish.”

2.4. \textsc{need}-passives and Other Low-Frequency Features

\textsc{need}-passives are described in sociolinguistic and dialectological literature as occurring infrequently in natural-language corpora. Murray, Frazer, and Simon (1996:258), for instance, relied on conscious judgments from respondents on the acceptability
of NEED+ED sentences as “a pragmatic decision based on the great difficulty we had in eliciting large quantities of information about [NEED+ED] through more traditional atlas-type methods or through relatively brief periods of free conversation.”

Illustrative of the low frequency of NEED-passives are the sociolinguistic interviews I conducted in Kansas City (Strelluf 2018), where sixteen of fifty Kansas Citians indicated they could use the sentence, “The car needs washed.” However, in thirty hours of casual speech during these interviews, there were no occurrences of NEED+ED. NEED+ING also never occurred, and there were just eight tokens of NEED+TO. Interviews conducted for The Scots syntax atlas (Smith et al. 2019) show that a speech corpus must be massive to generate just a small set of NEED-passives: 281 interviews conducted with 562 participants yielded twenty-seven instances of NEED+ED, eighteen NEED+TO, and three NEED+ING (E Jamieson, personal communication, 16-20 September 2019). Transitive-NEED+ING also occurs infrequently in speech. In a systematic survey of the 10-million word spoken component of the British National Corpus, De Smet (2013:84) found only eight instances. While it is inherently difficult to quantify exactly how rarely a feature must occur to be “low frequency,” it is qualitatively clear that very large corpora of spoken English generate very small counts of NEED-passives.

The low frequency of NEED-passives presents a fundamental challenge to quantitative approaches to the study of language variation and change. Labov (2006:32) set out the principle for linguistic variables that “the most useful items are those that are high in frequency, have a certain immunity from conscious suppression, are integral units of larger structures, and may be easily quantified on a linear scale.” While variationist methodologies have expanded and diversified tremendously since Labov’s foundational work in the 1960s, gathering a large sample of a variable and quantifying its occurrence or non-occurrence remains at the heart of Labovian sociolinguistics. The preference for structurally obligatory, high-frequency features is especially reflected in the central position of phonetic and phonological variables in variationist sociolinguistics but is also reflected in the variables that have been selected for morphosyntactic (e.g., Tagliamonte 2011:206-241) and discourse-pragmatic (e.g, Pichler 2016) analyses.

When a feature occurs infrequently and non-obligatorily (i.e., a speaker may elect to utter a NEED-passive to fill a discursive need, but NEED-passives are not syntactically required in any given utterance), quantifying its occurrence or non-occurrence does not yield meaningful analyses. As Murray and Simon (2002:34) argued in their study of the AEP LIKE+ED, non-use of a low-frequency feature “means only that an informant has not used it yet; the construction may appear in the next sentence or […] never.”

Accordingly, most studies of NEED-passives (specifically, NEED+ED) have followed the logic cited above in Murray, Frazer, and Simon (1996:258) and have gathered data by surveying consciously elicited judgments of grammatical acceptability. Examples include studies in sociolinguistics and dialectology (e.g., Labov, Ash & Boberg 2006: 293-296; Murray & Simon 2006:20; Bloomquist 2009; Maher and Wood 2011; Hickey 2018), as well as micro-syntax (e.g., Tenny 1998; Edelstein 2014—though Edelstein validates a number of claims about AEP constructions through web searches).
However, studies like Murray, Frazer, and Simon (1996; see also Murray & Simon 1999, 2002) and Labov, Ash, and Boberg (2006:293-296) that relied on consciously elicited judgments of the grammaticality of NEED+ED also warned that such judgments may be unreliable. Murray, Frazer, and Simon (1996:266) noted that “users of the construction often incorporate it into their language so unselfconsciously that some of them actually deny using it, then do use it only moments later without realizing they have done so.” Labov, Ash, and Boberg (2006:293-296) likened NEED+ED to another low-frequency feature, “positive anymore”—where the adverb anymore is used in positive-polarity clauses such as, “It’s real hard to find a good job anymore”—and warn that their map of elicited acceptability judgments must be interpreted with caution because speakers may not accurately recognize that they use the construction. Labov, Ash, and Boberg (2006:293) noted, “since it is not stigmatized overtly, and it is widely used by all social classes in speech, it is not yet clear why these intuitive responses differ so widely from practice.” Youmans (1986:71), also writing about positive anymore, attributed the unreliability of judgments to rarity: “[e]vidently, low-frequency phenomena such as positive anymore can be heard for years without registering on a listener’s consciousness.”

This places research on NEED+ED, on NEED-passives more generally, and potentially on other low-frequency features in a paradox. Because they occur infrequently in natural-language corpora, these features may be examined through the elicitation of conscious judgments. However, (perhaps because they occur infrequently) conscious judgments of these features may be unreliable (see also Strelluf 2019:321).

The present work is therefore undergirded by an interest in finding ways to work with NEED-passives and other low-frequency features that avoid this paradox. Narrowly, my approach here is to collect so much naturalistic language that it is possible to study NEED-passives according to core Labovian approaches that work for high-frequency variables and, in doing so, shed new light on a variable that has generated sociolinguistic, dialectological, and micro-syntactic interest. More broadly, at the heart of this interest is an observation that, because tools for the quantitative study of language variation and change are especially suited to higher-frequency features, theories of language variation and change have been built from datasets of higher-frequency features. It is not inherently the case that explanations and predictions built from high-frequency features will scale down to explain and predict the behavior of low-frequency features. Citations above from Youmans (1986), Murray, Frazer, and Simon (1996), and Labov, Ash, and Boberg (2006), for instance, described the surprising invisibility of NEED+ED and positive anymore to sociolinguistic monitoring and self-evaluation.

Strelluf (2020:129) further pointed out that NEED+ED “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,” and suggested that this endurance is not easily accounted for in variationist models of dialect contact and leveling. For instance, Trudgill’s (2004) influential model of new dialect formation, based especially on phonetic and phonological data from New Zealand, pointed to the primacy of majority forms in
determining which variant among several in competition will be selected for a new language variety. If the frequency of features like need-passives in spoken corpora reflects their frequency in language users’ interactions, it is unclear how the concept of a “majority form” might apply in a dialect contact situation. In a multilingual and multidialectal space like western Pennsylvania in the late 1700s or early 1800s, how would language users (particularly children acquiring language) have cognitively processed any need-passive construction as a majority form when they might have gone through huge stretches of language without encountering a need-passive? How would need-passives continue to be maintained in an area over generations as a trace of Ulster migrations?

These questions become more pointed under Edelstein’s (2014) syntactic analysis of the AEP as a syntactically unique structure, as this would require language users to maintain an idiosyncratic derivational operation just for a small set of matrix verbs to use for an apparently rare discursive requirement. High-frequency features would seem to be better suited to such idiosyncratic syntax than low-frequency features. High frequency has been extensively documented as a force for maintaining irregular morphosyntactic features (e.g., Bybee & Thompson 1997; Corbett, Hippisley, Brown & Marriott 2001) and for driving grammaticalization (e.g., Bybee 2007:269-357) (see Bybee & Hopper 2001 for discussion). Intuitively, Edelstein’s (2014) proposed syntax for the AEP would require that the construction either occur frequently enough to be subject to Bybee and Thompson’s (1997) “Conserving Effect” to maintain it in grammars, or frequently enough that matrix verb + AspP constructions might be reanalyzed as a particular type of constituent. There’s not an obvious mechanism for the maintenance or emergence of a novel syntax for a low-frequency feature in these models.

Focused attention to bring low-frequency features into the fold of quantitative approaches to language variation and change will inform the extent to which current theories account for low-frequency features. Such attention may explain surprising behaviors in low-frequency features and may contribute more broadly to theories and knowledge of language and the language faculty. Ultimately, theories of language are better if they describe or are confirmed to describe features regardless of frequency.

3. Methods

Strelluf (2019, 2020) demonstrated the usefulness of Twitter for building production-based corpora of low-frequency features. Because users post so much speech-like text to Twitter, rare features occur in sufficient numbers to allow quantitative analysis (see also Eisenstein, O’Connor, Smith & Xing 2012; Jones 2015; Pavalanathan & Eisenstein 2015; Squires 2016a, b; Eisenstein 2017; Grieve, Nini & Guo 2018).

I used the twitteR package (Gentry 2015) for R (R Core Team 2020) to sample tweets that contained a form of the word need.5 I sampled tweets daily between July 5 and September 4, 2018 from twenty US cities, seventeen UK cities, and thirteen other cities in countries with large English-speaking populations. I use the label “world” as a
shorthand for cities in the sample that are not in the United States or United Kingdom, rather than in the more thoughtful sense of scholarship of World Englishes (cf. works collected in Kachru, Kachru & Nelson 2006). Varieties included under this world label in the present study include, in the terminology of Kachru’s “Three Circles Model” (e.g., Kachru 1985), “Inner Circle” varieties of Canada, Ireland, and Oceania, where English is codified as a first language for most speakers, as well as “Outer Circle” varieties of Africa, Asia, and South America, where English is an “institutionalized additional language” (Kachru 2005:14). The full list of sampled cities appears in Appendix 1.

Tweets associated with a geographical area in Twitter data are not assuredly representative of that area in the way that dialectologists traditionally require. Twitter’s public search interface samples tweets based on the physical location of the device that tweeted the message, and also samples on the basis of locations that users enter in their profiles. As such, an unknowable number of tweets will be associated with locations where an author did not acquire language as a child. Nevertheless, a working assumption is that datasets built from Twitter are so large that good data will suppress noise. Researchers have demonstrated that geographically associated tweets can generate robust dialect maps (e.g., Eisenstein, O’Connor, Smith & Xing 2012; Jones 2015; Pavalanathan & Eisenstein 2015). I therefore report locations for tweets but acknowledge the inherent noisiness of the data.

I did not collect any social information on authors. Importantly, this means that people whose tweets were sampled in an area will be captured under the same areal label, even though conventional sociolinguistic or dialectological studies might treat them as being speakers of different sociolects or ethnolects (see Strelluf [2020:127] for discussion of NEED+ED in African American Language). As before, the single social variable of “location” is noisy in this study.

The two-month Twitter scrape resulted in an initial pool of more than 3.6 million tweets. I tagged all words in all tweets for part-of-speech with the TwitIE scripts (Derczynski, Maynard, Aswani & Bontcheva 2013:21; Bontcheva et al. 2013; Derczynski, Ritter, Clark & Bontcheva 2013). Tagging procedures were detailed in Strelluf (2020), but as I developed methods for this project, taggers nearly always failed to tag NEED+ED—apparently disallowing the possibility that a participle could immediately follow a NEED lexeme, and usually tagging passive complements as nouns. In this project, I prevented the tagger from coding NEED lexemes as verbs, which caused it to tag anything that looked like a participle after NEED as a verb. The broader methodological observation, though, is that low-frequency features may naturally pose challenges for taggers—low-frequency features are unlikely to occur in a training corpus, or they occur so infrequently that algorithms will assign low probability that a given occurrence in a test corpus is that feature.

After tagging, I extracted all tweets where NEED was followed immediately by a word tagged as a past participle, gerund-participle, or to be and a past participle, or where any of these constructions occurred with an intervening adverb or negative particle. Constructions with intervening adverbs or negation were not included in the dataset reported in Strelluf (2020). In both studies I used aggressive filters to drop
tweets from the datasets where formatting oddities or problematic characters created the potential for errors to be read into R, with the effect that many tweets that were included in the dataset for Strelluf (2020) were dropped from the new dataset. As such, while Strelluf (2020) and the present study pull from the same initial pool of tweets, the studies do not contain all the same tweets.

The tagging process resulted in a corpus of 44,290 tweets tagged as NEED+TO, 14,496 NEED+ING, and 6984 NEED+ED. I manually checked these 65,770 tweets for tagging errors. Routine errors included nominal NEED being tagged as a verb as in (10)-(12), modifiers after NEED being tagged as verbs as in (13)-(15), and other instances of text after NEED being tagged as a verb due to misspellings and unconventional formatting as in (16) and (17) or not appearing in TwitIE’s dictionary (18).

(10) People have unique needs to be met after every disaster. (New Delhi; tagged as NEED+TO)
(11) I shall address my voters needs including transport. (Liverpool; tagged as NEED+ING)
(12) Boy with special needs killed. (Pittsburgh; tagged as NEED+ED)
(13) This needs to be required reading for parents. (Phoenix; tagged as NEED+TO)
(14) He doesn’t need running shoes. (Phoenix; tagged as NEED+ING)
(15) I need sprinkled donut. (Minneapolis; tagged as NEED+ED)
(16) He need to be STRAAAAAAAAAAAAAAIT with his level of interest. (Boston; tagged as NEED+TO)
(17) I need atleast two. (London; tagged as NEED+ED)
(18) What you need melatonin for? (Columbus; tagged as NEED+ING)

Of methodological note, NEED+TO and NEED+ING were less sensitive than NEED+ED to tagging errors resulting from spelling. In (19) and (20), the tagger correctly interpreted novel spellings of educated and re-negotiating. As such, the tagger seemed to be likely to miss occurrences of NEED+ED where it would identify NEED+TO and NEED+ING. Researchers of low-frequency features should interrogate their datasets for similar imbalances.

(19) People need to be eductd. (Islamabad; correctly tagged as NEED+TO)
(20) They need renegotiating because they are part of deals. (Manchester; correctly tagged as NEED+ING)

A few tweets were correctly tagged as containing a NEED-passive but were tagged for the wrong construction. This occurred most frequently when NEED was followed by the intensifier fucking (21), usually resulting in NEED+ED tweets being erroneously tagged as NEED+ING. Such cases were re-coded to the correct NEED-passive.

(21) Whoever evaluates quality needs fucking sacked. (Glasgow; tagged as NEED+ING)
A small set of interesting but irrelevant NEED constructions were erroneously tagged as containing NEED+ED or NEED+ING. In (22) and (23), need means “never.” Elsewhere, NEED occasionally took a non-passive non-finite complement headed by a progressive verb (24) or plain form (25). Such constructions were excluded.

(22) We are need drinking ever again. (Edinburgh; tagged as NEED+ING)
(23) I’ve need had a pigeon poop on me. (Minneapolis; tagged as NEED+ED)
(24) That needs going in the rubbish. (Manchester; tagged as NEED+ING)
(25) He need put more swing in the hips. (Philadelphia; tagged as NEED+ED)

The NEED+ING sample was complicated by the fact that nouns can end in -ing. Tweets tagged as NEED+ING were only retained if I could felicitously rephrase them as NEED+TO, and I regularly checked my intuitions by confirming that the past participle form of the verb also occurred as a passive complement in either the NEED+TO or NEED+ED datasets. A small subset of complements—counseling, financing, funding, and healing—occurred relatively frequently in the NEED+ING dataset and, while they occurred as passive verbal complements to NEED+TO, almost never occurred as complements in the NEED+ED dataset. On the possibility that these were not comparable to NEED+ED, I excluded all tweets in the corpus of NEED+ING tweets with these four complements.

NEED+ED complements were also checked to confirm that they could be rephrased as NEED+TO, and that they occurred as complements with NEED+TO or NEED+ING. This resulted in exclusions, but there were not systemic errors in NEED+ED for specific lexemes.

The NEED+TO sample included tweets where the complement was a participial adjective rather than a verb. Because NEED+ED prohibits adjectival complements (see discussion of Tenny [1998] and Edelstein [2014] above), NEED+TO tweets with adjectival complements were excluded. Among the most frequently excluded complements were done with, concerned, gone, lit (“drunk,” “fun”), married, and worried.

I did not intend to sample transitive-NEED constructions. However, as examples (4)-(6) illustrate, when the passive object complement is dislocated leftward, the matrix verb NEED and participle verb in the complement end up next to each other. During tagging, these look the same as NEED+TO, NEED+ING, and NEED+ED. 357 constructions were erroneously tagged as NEED+ED and recoded to transitive-NEED+ED, and sixty-four were tagged as NEED+ING and recoded to transitive-NEED+ING. These were excluded from the corpus of NEED-passives but will be examined opportunistically as their own dataset. There were only three occurrences of transitive-NEED+TO; these are not analyzed.

I deleted any tweet that was sent as identical or nearly identical text from a single handle (i.e., cases where a single author tweeted basically the same text more than once). I did not sample tweets that Twitter classified as retweets. However, in cases where more than one author tweeted very similar text as an original message from their own handle, I kept these in the dataset. I made this decision because an author’s re-broadcasting of a tweet creates ownership over the message that is not present in a retweet (e.g., it appears...
on Twitter as a message from their handle), and because it would have been possible for authors to edit a NEED-passive construction if they had objected to it.

Finally, 650 tweets were excluded because I could not interpret them.

These procedures left a corpus of 41,668 instances of NEED+TO, 9935 NEED+ING, and 3232 NEED+ED. A comparison of the final numbers of tweets containing each NEED-passive included after error-checking against the pool that was initially sampled shows the degree to which the effectiveness of tagging procedures differed across NEED-passives: 94 percent of NEED+TO tweets were retained versus 68 percent of NEED+ING versus 46 percent of NEED+ED. I attribute these differences in the relative success of tagging NEED-passives to the role that frequency plays in training taggers and note this as an additional practical challenge to studying low-frequency features.

4. Results

4.1. NEED-passive Proportions in Global Englishes

Appendix 1 reports counts for each form of NEED-passive in all fifty cities sampled, as well as how frequently each NEED-passive occurs as a proportion of all NEED-passives in each city. These proportions are represented visually in Figure 4, which reflects an inductive approach to allow the NEED-passive proportions to organize the cities into a single intuitive view. Cities where NEED+ED occurs more frequently than NEED+ING are sorted in descending order of their NEED+ED proportion. All other cities are sorted in ascending order of their NEED+ING proportion.

NEED+TO proportions in Figure 4 show clearly that NEED+TO is the majority NEED-passive in global Englishes, accounting for more than 50 percent of NEED-passives in forty-three of fifty cities. There is an obvious pattern among the cities with the lowest NEED+TO proportions: the sixteen lowest NEED+TO proportions are cities in the United Kingdom. The cities with greatest NEED+TO proportions include Outer Circle varieties of New Delhi, Islamabad, and Cape Town, as well as most of the US varieties outside the Midland. Generally speaking, NEED+TO is the global default NEED-passive construction everywhere except the United Kingdom and US Midland.

The right side of Figure 4 creates a strong visual impression of inter-variety differences in NEED+ING proportions. The thirteen highest NEED+ING proportions belong to cities in England and the Welsh capital, Cardiff. Indeed, all the English cities in the sample land in this cluster. The greatest NEED+ING proportions belong to Liverpool, Leeds, and Manchester, all in the English North, and these are followed by Nottingham, Northampton, and Birmingham in the English Midlands. In contrast to the English and Welsh cities, NEED+ING occurs uniformly as a low proportion of NEED-passives in the United States. Seattle’s NEED+ING proportion of 11.4 is the greatest among all twenty US cities. As with NEED+TO, Cape Town, Islamabad, and New Delhi align with US cities in NEED+ING proportions. Other cities, including UK varieties of Scotland and Belfast, cluster in a range of NEED+ING proportions from 10.4 in Toronto to 21.3 in Auckland.
NEED+ED occurs as a tiny fraction of NEED-passives in most cities. The left side of Figure 4 highlights NEED+ED as a feature of Belfast, all three Scottish cities, and the US Midland cities of Pittsburgh, Columbus, and Indianapolis. NEED+ED proportions decrease down to Kansas City in the western range of the US Midland, after which point all cities have greater NEED+ING than NEED+ED proportions.

The US city Cleveland, which is classified as part of the North in current American dialectology on the basis of phonetic and phonological analyses (e.g., Labov, Ash & Boberg 2006:194), joins geographically nearby Columbus and Pittsburgh in having a relatively high proportion of NEED+ED. On the right side of Figure 4, Newcastle is a visually striking outlier among the English cities as increased NEED+ED displaces NEED+TO, while the relatively high proportion of NEED+ING that is typical of England is also maintained. Consistent with Newcastle’s geographic position between the English and Scottish cities in the sample and its deep historical and cultural connections to both the English North and Scotland, Newcastle is unique among the cities in this sample for featuring both the high proportion of NEED+ING that is associated with England and the high proportion of NEED+ED that is associated with Scotland.

The proportions reported in Figure 4 confirm previous characterizations of NEED+ED as a regional grammatical feature of Belfast, Scotland, Newcastle, and the US Midland (e.g., Strelluf 2020). They also reveal variation across Englishes in NEED+ING, with these constructions occurring in greater proportions in the English North particularly
and in England and Wales more generally. They show NEED+TO to be overwhelmingly preferred in a range of Englishes that includes Cape Town, Islamabad, and New Delhi, as well as most US cities.

Three-way variation in Englishes among NEED+TO, NEED+ING, and NEED+ED is confirmed by cluster analysis. Cluster analysis algorithms chunk observations into an analyst-specified number of groups in order to achieve the greatest possible similarity among observations within each group. Figure 5 shows an output of a K-means cluster analysis created with kmeans() in R (R Core Team 2020) using the default Hartigan-Wong algorithm (Hartigan 1975; Hartigan & Wong 1979), which assigns observations to clusters so that the sum of squares between the observations and the center point of their assigned cluster is minimized.

K-means clustering requires normalized data, so I scaled the NEED-passive proportions in Appendix 1 around a mean of 0 and standard deviation of 1 using R’s built-in scale() function. The factoextra package (Kassambara & Mundt 2020) in R provides three functions for estimating the optimal number of groups to enter in a cluster analysis (“elbow method” [e.g., Thorndike 1953], “average silhouette” [e.g., Rousseeuw 1987], and “gap statistic” [e.g., Tibshirani, Walther & Hastie 2001]). All three functions converged on three groups as optimal, so Figure 5 shows a K-means cluster analysis based on three groups, and visualized with fviz_cluster() from factoextra. Since there are three NEED-passive variables, the function creates a two-dimensional plot by performing a principal component analysis and then plotting according to the first two principal components. In Figure 5, NEED+ING corresponds to the x-axis and NEED+ED the y-axis.

Figure 5. Cluster Analysis of NEED-passives in All Cities
Figure 5 plots three clusters that generally reflect the qualitative analysis of need-passive proportions. In the bottom-left, a cluster is formed of Cardiff and all English cities except London. At the top of the figure are Belfast, the three Scottish cities, and Pittsburgh and Columbus. The cluster at bottom-right includes London, all world Englishes, and all other US Englishes—including cities with relatively high need+ed proportions in Figure 4 such as Cleveland, Indianapolis, and Kansas City. The English and Welsh cluster corresponds to the area of greatest need+ing proportions. The Scottish, Belfast, and eastern-most US Midland cluster corresponds to the area of greatest need+ed proportions. Newcastle reaches up toward this cluster but is still grouped with other English cities in the need+ing cluster. The final group includes all other varieties.

This clustering is indicative of three distinct need-passive regions. Need+ing is a syntactic variant of England and Wales in the same way that need+ed is a variant of Northern Ireland, Scotland, and parts of the US Midland. Englishes elsewhere coalesce around need+to.

The patterning of Englishes according to these three constructions, in some cases, offers fascinating reflections of historical connections among varieties. The cline of need+ed proportions among US Midland cities follows Strelluf (2020) in showing proportions reducing steadily from east to west among Pittsburgh, Columbus, Indianapolis, and Kansas City, suggestive of a westward diffusion of need+ed across the Midland along migration routes of white settlers of Ulster Irish descent in the 1800s (see Montgomery 1991, 1997). Analogously, among the non-British varieties nearest to the need+ing end of the continuum are Inner Circle varieties with obvious historical ties to English settlement in Auckland, Dublin, and Sydney (e.g., Gordon et al. 2004; Hickey 2007; Cox & Palethorpe 2007). Canadian varieties in Toronto and Vancouver land on the continuum in Figure 4 to the English side of the US cities, suggestive of other linguistic features where Canadian Englishes are generally similar to Englishes of the northern and western United States, but still distinct and maintaining some Britishisms (Chambers 1995; Boberg 2010). Newcastle’s unique status as a city with high proportions of both need+ing and need+ed aligns not only with the city’s geographical position, but also with linguistic roots tracing back, as an anonymous reviewer pointed out, more than a millennium to Anglo-Saxon Northumbria.

On the other hand, the patterning of other cities cannot be explained as tidily. Manila, an Outer Circle variety with colonial roots in American English (Lim 2012), appears to orient toward British need+ing rather than the general US dispreference for anything but need+to. Speculatively, Cape Town’s alignment with Islamabad and New Delhi could reflect the large populations of Indians in South Africa (Mesthrie 1992), but there is not a readily obvious explanation for the Outer Circle Englishes of India and Pakistan avoiding need+ing while other post-colonial varieties in Georgetown, Hong Kong, Lagos, and Singapore have proportions closer to Inner Circle varieties in Australia and New Zealand.
The difficulty explaining patterns in Englishes outside the United States and United Kingdom, however, does not detract from the strength of the fundamental observation that there are three patterns of NEED-passive grammars. NEED+ED is a regional syntactic feature of Northern Ireland, Scotland, and the US Midland. These NEED+ED grammars are further differentiated by strength of preference for NEED+ED relative to other NEED-passives. NEED+ING is more common in Englishes than NEED+ED, but it is actually also a regional syntactic feature of England and Wales. Again, there is proportional variation within the NEED+ING grammars, with the feature being especially concentrated in the English North. Elsewhere, especially in North America and (speculatively) post-colonial Englishes associated with the British in India, NEED+TO is strongly preferred, to the point that both NEED+ED and NEED+ING might be regarded as marginal features.

4.2. Transitive-NEED

Appendix 2 lists the counts of transitive-NEED+ED and transitive-NEED+ING for each city, as well as the proportions of each construction that counts represent. It is immediately clear in Appendix 2 that transitive-NEED+ING is a construction of England and Wales. Outside England and Wales, only Hong Kong has more transitive-NEED+ING than transitive-NEED+ED, resulting from a single tweet shown in (26).

(26) These are ten questions you need answering before you apply. (Hong Kong)

By contrast, in Cardiff and every English city except London and Newcastle, at least half of NEED-transitives are transitive-NEED+ING. In most cities, counts are quite small, but cases like Manchester, where thirteen of fifteen transitive constructions are transitive-NEED+ING, give credence to a pattern. Pearson’s product-moment correlation tests show that cities’ proportions of NEED+ING and transitive-NEED+ING are strongly linked ($r = 0.834; p < .001$). As a city’s proportion of NEED+ING increases, so does its proportion of transitive-NEED+ING. NEED+ED does not significantly predict transitive-NEED+ED ($r = 0.177; p = .245$).

The unintended sample of transitive-NEED is small and must be interpreted with caution. Data indicate, however, that in England and Wales NEED more frequently uses a gerund-participle in forming both NEED-passives and transitive-NEED. Other varieties mostly reserve the past participle for transitive-NEED. This suggests that Huddleston and Pullum’s (2002:1245) description of transitive-NEED+ED as a regionally restricted construction and transitive-NEED+ING as a general feature of Englishes is incorrect. More fundamentally, though, transitive-NEED adds an additional layer of complexity to intra-English differences in the syntax of NEED constructions.
4.3. Syntactic Observations

This section applies syntactic tests from Tenny (1998) and Edelstein (2014) to tweets. In the case of NEED+ED, these tests check whether productions in Twitter align with Tenny’s (1998) and Edelstein’s (2014) (primarily) judgment-based data. In the case of NEED+ING, the tests check whether NEED+ING follows a similar set of syntactic constraints to NEED+ED.

The first two tests I apply were used by Tenny (1998) and Edelstein (2014) to show that NEED+ED complements are always verbal rather than adjectival. Because NEED+TO complements can be either adjectival or verbal, this analysis indicates that NEED+TO and NEED+ED result from different derivations.

Tenny (1998:592) and Edelstein (2014:261-262) noted that NEED+ED complements allow purposive by-phrase adjuncts, which force a verbal reading. Example (27) shows one of forty-six instances of NEED+ED taking a purposive by-phrase in the dataset. Example (28) shows the same for NEED+ING, which occurred in thirty-nine tweets. In the case of NEED+ING, the purposive by-phrase forces a verbal rather than nominal reading (see Huddleston & Pullum [2002:1200] for this analysis of by-phrases in gerund-participle non-finite clauses).

(27) Trash bags need picked up by KCMO Waste Department. (Kansas City)
(28) That lad needs teaching a lesson by you. (Manchester)

Tenny (1998:593) and Edelstein (2014:260) claim that NEED+ED cannot take a passive complement with a non-reversive un-prefix. Because non-reversive un- can only affix to adjectives, its non-occurrence in the AEP provides further evidence that NEED+ED passive complements are always verbal. Tweets support this analysis. There are no NEED+ED or NEED+ING tweets where the passive complement has a non-reversive un-prefix. There are four instances of reversive un-prefixes on NEED+ED complements and twenty-four instances with NEED+ING, exemplified in (29) and (30).

(29) Just like Pereira would unlock Pogba, now Lingard needs unlocked too? (Liverpool)
(30) Looks like Klopp needs unlocking. (Liverpool)

The presence of purposive by-phrase adjuncts to passive complements and of passive complements with non-reversive un-prefixes confirms, specifically in the cases of the tweets where they occur, that the passive complements are verbal. To be clear, these data do not show that passive complements to either NEED-passive must be verbal. However, NEED+ED data fail to contradict Tenny’s (1998) and Edelstein’s (2014) claims that NEED+ED allows only verbal complements. In doing so, NEED-passive productions on Twitter offer no challenge to the conclusions Tenny (1998) and Edelstein (2014) each reach from acceptability judgments about NEED+ED syntax. Indeed, the extension of Tenny’s (1998) and Edelstein’s (2014) tests to NEED+ING suggests that it would be
valuable to test the acceptability of NEED+ING constructions that force verbal readings, as in (31), against constructions that allow or require the NEED+ING complement to be nominal, as in (32) and (33).

(31) Education needs overhauling by experts. (Lagos; necessarily verbal)
(32) Education needs overhauling. (Constructed; ambiguously verbal or nominal)
(33) Education needs some serious overhauling. (Constructed; necessarily nominal)

The second two tests I apply were used by Edelstein (2014) to support the analysis of a novel derivational structure for the AEP (see Figure 1), which differed from that of the standard embedded passive (Figure 2) or concealed passive (Figure 3). The tests indicated a closer syntactic relationship between an AEP matrix verb and its passive complement than would exist under derivations of a non-finite clause being taken as a complement to a matrix clause. Edelstein’s (2014) proposal that, in the AEP, the matrix verb takes an AspP as a complement rather than a TP or complementizer phrase provides this closer syntactic relationship.

Responses to Edelstein’s (2014:265-266) survey of acceptability judgments showed that adverbial interruptions between an AEP matrix verb and its passive complement were dispreferred. NEED-passives with intervening adverbs are exemplified in (34)-(36).

(34) This needs to be seriously publicized. (St. Louis)
(35) Zara needs absolutely booting off this series. (Manchester)
(36) These horrible places need permanently shut down. (Pittsburgh)

In the Twitter corpus, adverbs ending in -ly sit between the matrix verb and passive complement nearly twice as frequently in NEED+TO ($N = 648$; 1.6 percent of NEED+TO tweets) as in NEED+ED ($N = 24$; 0.7 percent). The dispreference for adverbial interruption is even greater in the case of NEED+ING ($N = 29$; 0.3 percent of NEED+ING tweets), which is interrupted by an -ly adverb at one-fifth the rate of NEED+TO.

Edelstein’s (2014:264) respondents also rejected AEP constructions where negation intervened between the matrix verb and the passive complement, as in the disallowed sentence in (37). NEED+TO does allow negation between the matrix verb and passivized verb, as in (38).

(38) I need to not be questioned. (Cape Town)

In the Twitter corpus, neither NEED+ED nor NEED+ING occurs with negation between the matrix verb and passive complement. By contrast, NEED+TO is interrupted by not or never in fifty tweets.

Edelstein’s (2014) indicators of the validity of her unique AEP derivation are upheld for NEED+ED in tweets. Moreover, these indicators appear to be present for
NEED+ING too. For some language users, at least, the passive complement in
NEED+ING appears to hold a tighter syntactic relationship to the matrix verb than does
the passive complement in NEED+TO. This mirrors Edelstein’s (2014) analysis of
NEED+ED syntax.

It seems plausible that Edelstein’s (2014) novel derivational structure for AEPs can
then extend to the derivation of concealed passives. This suggestion is bolstered by the
broader observation that the AEP and concealed passive are formed from the same set
of matrix verbs. As such, results point toward a reanalysis of the syntax of concealed
passives as derivationally identical to AEPs. Following Edelstein’s (2014:265; Figure 1
above) analysis of the AEP, in the concealed passive, the matrix verb would directly
select an AspP complement, and that AspP would assign passive morphology to the
verb in the complement. The AEP and concealed passive would differ only in the
participle form assigned by AspP.

This analysis reconceptualizes previous framings of the syntax of NEED-passives.
Edelstein (2014) reflected the traditional treatment of NEED+ED as a regionally
constrained alternative to NEED+TO in using the labels “embedded passive” for
NEED+TO and “alternative embedded passive” for NEED+ED, as well as retaining
Huddleston and Pullum’s (2002) label of the “concealed passive” for NEED+ING. If
NEED+ING is actually formed from the same derivation as NEED+ED, then labels for
these constructions should reflect their shared syntax—i.e., both NEED+ING and
NEED+ED are either AEPs or concealed passives. The present analysis also addresses
Murray and Simon’s (1999:158) speculation in the context of passives formed with
WANT that “there must be two rules, one in the underlying grammar of [WANT+ED]
users, the other in the underlying grammar of [WANT+ING] users, that block the
formation (and hence the acceptance) of the alternate construction.” In this revised
analysis, there is one rule in the underlying grammar and users differ only super-
ficially in participle morphology.

5. Discussion and Conclusion

This study has revealed NEED-passives to be a complex system of inter-variety variation
in Englishes around the world. It has recast NEED+ING as a regional syntactic variant
that is distinctive to England and Wales in the same way that NEED+ED is distinctive to
Northern Ireland, Scotland, and the US Midland. It has also united most Englishes in
North America, India, Pakistan, and South Africa in an overwhelming preference for
NEED+TO.

These results call for reanalysis of previous research on NEED-passives, especially in
the United States. They also open space for new investigations. NEED+ING should be
brought squarely into the fold of British dialectology. Newcastle bears examination as a
space where all three NEED-passives appear to be on a fairly equal footing. And the
limited data in this study for Englishes outside the United States and United Kingdom
suggest that NEED-passives should be examined more broadly as a variable across global
Englishes.
The observation that NEED+ING has escaped notice as a regional variant has a parallel in Murray, Frazer, and Simon’s (1996:255-256) puzzlement over why NEED+ED had received “remarkably little attention” from linguists. While NEED+ED seems to have been mostly invisible to linguists before Stabley’s (1959) note in the miscellany of American Speech, NEED+ING may have been hiding in plain sight. Low frequency is likely to blame. Citations of Youmans (1986), Murray, Frazer, and Simon (1996), Murray and Simon (1999, 2002), and Labov, Ash, and Boberg (2006:293-296) have posited a role for low frequency in leaving some variables invisible to social evaluation or conscious recognition. NEED+ING may be an even more complicated case, because recognizing its nature as a regional variant requires a fine analysis of frequency to distinguish between the low levels of NEED+ING that occur in all Englishes and the elevated proportions of NEED+ING in Britain. Of course, fine analysis of frequency is exactly the sort of analysis that will be blocked by low-frequency features.

The effort to overcome the methodological problem of low frequency has resulted not only in the identification of previously unrecognized variability among NEED-passives, but also in syntactic reanalysis. This reanalysis may point the way toward resolving some of the mysterious characteristics of NEED-passives as low-frequency features.

For instance, the invisibility of NEED+ED to conscious evaluation noted in Murray, Frazer, and Simon (1996) and Labov, Ash, and Boberg (2006:293-296) may be less surprising if NEED+ED and NEED+ING are derived from the same underlying syntactic operation. A language user whose NEED-passive grammar assigns a gerund-participle might use that derivation to process (or rescue) an utterance that differs only in containing a past participle. There is unintended support for this suggestion in psycholinguistic studies by Kaschak and Glenberg (2004) and Kaschak (2006), which showed that English speakers who were unfamiliar with NEED+ED could be exposed to it, and then rapidly and accurately generalize it to other matrix verbs. It is possible that participants’ rapid and accurate acquisition of NEED+ED did not reflect a general cognitive ability, but rather a specific fact that the syntax for NEED+ED was already part of their grammar as NEED+ING, so participants only had to learn to substitute a different participle form. If this explanation is correct, speech communities like Newcastle that use both NEED+ING and NEED+ED (or speakers who use both forms) are unsurprising. Rather than one grammar “blocking” the other, as Murray and Simon (1999:158) speculated, it would be relatively straightforward for a mental grammar to allow both participle forms to mark passive morphology.

Perhaps a similar mechanism could factor into the maintenance of NEED+ED or NEED+ING in speech communities across long stretches of time. If NEED+ED and NEED+ING share the same syntax, then combined exposure to them would potentially increase the actual exposure to the underlying derivation. In other words, perhaps someone could be exposed to the syntax of NEED+ED by being exposed to NEED+ING, and vice versa. NEED-transitives could play a role here, too. NEED+ING and transitive-
NEED+ING can, in principle, both be generated by the syntax Edelstein (2014) proposes for the AEP. Indeed, Huddleston and Pullum’s (2002:1206) description suggests that NEED+ING and transitive-NEED+ING both result from the embedding of a non-finite clause within a matrix clause, and just differ in whether the passive object raises all the way to the subject of the matrix verb. A shared syntax across NEED+ING, NEED+ED, transitive-NEED+ING, and transitive-NEED+ED might further reinforce the shared derivation.

On the other hand, co-variation between NEED+ING and transitive-NEED+ING may reflect more subtle differences in NEED-passive syntaxes across Englishes. Perhaps in England and Wales—where NEED+ING and transitive-NEED+ING are used as relatively high proportions of passivized constructions—speakers derive these constructions via the same derivational operation. However, in the Englishes where transitive-NEED+ED is the preferred NEED-transitive and where NEED+ING is produced more than NEED+ED, perhaps NEED+ING is not derived by the AEP syntax. In these grammars, the gerund-participle complement to NEED+ING might actually be nominal. This possibility would need to be tested by consciously elicited grammaticality judgments, as described in the context of sentences (31)-(33) in section 4. It would be confirmed through greater acceptability of sentences that force verbal readings of NEED+ING complements in England and Wales, while language users elsewhere in the world would prefer sentences that allow or force nominal readings.

If this speculation were borne out, it would limit the explanatory power of the shared-syntax account I have offered for NEED-passives. However, it would reveal a compelling new layer of variation in NEED-passives: that NEED+ING sentences uttered in different varieties of English, which look at surface level to be identical, might actually be derived from different derivational operations.

NEED-passives occupy a very small niche of English grammar. This study has revealed that the small space is a complex one, with a richer profile of variation across Englishes than has previously been recognized in dialectology, sociolinguistics, and micro-syntax, as well as in major grammars. I have argued that low-frequency features like NEED-passives are naturally subject to such mis- or under-analysis because linguists’ tools are not well-suited for studying them. While my approach in this paper was simply to collect enough naturalistic language to examine a low-frequency feature along the lines of higher-frequency features, I hope the approach will foster additional creativity and innovation in methods to study low-frequency features. The enriched account of NEED-passives provided by this work illustrates the possibilities for enriching descriptions and theories of English grammar, language variation and change, and the language faculty more generally through intensive attention to low-frequency features.
### Appendix 1

<table>
<thead>
<tr>
<th>Variety</th>
<th>City</th>
<th>Counts</th>
<th>Proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Need+TO</td>
<td>Need+ING</td>
</tr>
<tr>
<td>UK</td>
<td>Aberdeen</td>
<td>110</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Belfast</td>
<td>160</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Birmingham</td>
<td>589</td>
<td>461</td>
</tr>
<tr>
<td></td>
<td>Cardiff</td>
<td>432</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>Edinburgh</td>
<td>485</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>Glasgow</td>
<td>777</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Leeds</td>
<td>656</td>
<td>790</td>
</tr>
<tr>
<td></td>
<td>Liverpool</td>
<td>836</td>
<td>1137</td>
</tr>
<tr>
<td></td>
<td>London</td>
<td>1433</td>
<td>534</td>
</tr>
<tr>
<td></td>
<td>Manchester</td>
<td>1418</td>
<td>1584</td>
</tr>
<tr>
<td></td>
<td>Newcastle</td>
<td>529</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>Northampton</td>
<td>123</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>Norwich</td>
<td>214</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Nottingham</td>
<td>595</td>
<td>572</td>
</tr>
<tr>
<td></td>
<td>Peterborough</td>
<td>101</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Plymouth</td>
<td>95</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Southampton</td>
<td>247</td>
<td>168</td>
</tr>
<tr>
<td>US</td>
<td>Atlanta</td>
<td>762</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Birmingham</td>
<td>538</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Boston</td>
<td>1796</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>Chicago</td>
<td>1435</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>Cleveland</td>
<td>568</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Columbus</td>
<td>554</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Dallas</td>
<td>1487</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>Denver</td>
<td>681</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Detroit</td>
<td>666</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Indianapolis</td>
<td>937</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Kansas City</td>
<td>848</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Los Angeles</td>
<td>911</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Minneapolis</td>
<td>1076</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>969</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Philadelphia</td>
<td>1797</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>Phoenix</td>
<td>1838</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>Pittsburgh</td>
<td>1118</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>San Francisco</td>
<td>1539</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Seattle</td>
<td>2191</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td>St Louis</td>
<td>1097</td>
<td>81</td>
</tr>
</tbody>
</table>

(continued)
Appendix 2

<table>
<thead>
<tr>
<th>Variety</th>
<th>City</th>
<th>Counts</th>
<th>Proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Need+TO</td>
<td>Need+ING</td>
</tr>
<tr>
<td>World</td>
<td>Auckland</td>
<td>273</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Cape Town</td>
<td>471</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Dublin</td>
<td>1034</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>Georgetown</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>180</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Islamabad</td>
<td>387</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Lagos</td>
<td>576</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Manila</td>
<td>455</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>New Delhi</td>
<td>1640</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>510</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Sydney</td>
<td>1494</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>Toronto</td>
<td>1894</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>Vancouver</td>
<td>1142</td>
<td>176</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variety</th>
<th>City</th>
<th>Transitive-Need+ED</th>
<th>Transitive-Need+ING</th>
<th>Total</th>
<th>Transitive-Need+ED</th>
<th>Transitive-Need+ING</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Belfast</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>66.7</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>Birmingham</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>Cardiff</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>28.6</td>
<td>71.4</td>
</tr>
<tr>
<td></td>
<td>Edinburgh</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>85.7</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>Glasgow</td>
<td>12</td>
<td>1</td>
<td>13</td>
<td>92.3</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>Leeds</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Liverpool</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>25.0</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td>London</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>57.1</td>
<td>42.9</td>
</tr>
<tr>
<td></td>
<td>Manchester</td>
<td>2</td>
<td>13</td>
<td>15</td>
<td>13.3</td>
<td>86.7</td>
</tr>
<tr>
<td></td>
<td>Newcastle</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>87.5</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>Northampton</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Norwich</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>Nottingham</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>44.4</td>
<td>55.6</td>
</tr>
<tr>
<td></td>
<td>Peterborough</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>25.0</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td>Plymouth</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Southampton</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>50.0</td>
<td>50.0</td>
</tr>
</tbody>
</table>
Acknowledgments

This project has benefited tremendously from feedback and advice from editors Alexandra D’Arcy and Peter Grund and from three anonymous reviewers, as well as editorial support from Erika Grandstaff. I thank David Adger for patiently re-teaching me several points of derivational analysis, which became central to this account, and Claire Childs, whose intuitions about transitive-\textsc{need}+\textsc{ing} showed me gaps in my thinking. Finally, I’ve appreciated insights on \textsc{need}-passive variation shared by colleagues in connection with presentations at York and Warwick.

<table>
<thead>
<tr>
<th>Variety</th>
<th>City</th>
<th>Transitive-Need+ED</th>
<th>Transitive-Need+ING</th>
<th>Total</th>
<th>Transitive-Need+ED</th>
<th>Transitive-Need+ING</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>Atlanta</td>
<td>11</td>
<td>0</td>
<td>11</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Birmingham</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Boston</td>
<td>16</td>
<td>0</td>
<td>16</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Chicago</td>
<td>20</td>
<td>1</td>
<td>21</td>
<td>95.2</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>Cleveland</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Columbus</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Dallas</td>
<td>24</td>
<td>1</td>
<td>25</td>
<td>96.0</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Denver</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Detroit</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Indianapolis</td>
<td>14</td>
<td>0</td>
<td>14</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Kansas City</td>
<td>10</td>
<td>2</td>
<td>12</td>
<td>83.3</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>Los Angeles</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Minneapolis</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Philadelphia</td>
<td>24</td>
<td>3</td>
<td>27</td>
<td>88.9</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>Phoenix</td>
<td>16</td>
<td>0</td>
<td>16</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Pittsburgh</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>San Francisco</td>
<td>11</td>
<td>1</td>
<td>12</td>
<td>91.7</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>Seattle</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>St Louis</td>
<td>15</td>
<td>0</td>
<td>15</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>World</td>
<td>Auckland</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Cape Town</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Dublin</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>85.7</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Lagos</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>75.0</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>Sydney</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Toronto</td>
<td>22</td>
<td>2</td>
<td>24</td>
<td>91.7</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>Vancouver</td>
<td>16</td>
<td>1</td>
<td>17</td>
<td>94.1</td>
<td>5.9</td>
</tr>
</tbody>
</table>
Universities, Edisyn 9, ICLaVE 10, and ADS, especially E Jamieson, Elspeth Edelstein, Esther Asprey, Josef Fruehwald, and Raymond Hickey. I am responsible for shortcomings in the project.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Christopher Strelluf  
https://orcid.org/0000-0003-1539-1254

Notes
1. In presenting example tweets, I have deleted content that occurs before or after the construction under consideration. These reductions are meant to highlight the relevant syntax. Additionally, while I have left need lexemes, their passive complements, and any intervening words exactly as they appear in tweets, I have changed about 20 percent of words elsewhere in tweets. This procedure was recommended by Tatman (2018) to protect the anonymity of Twitter authors whose tweets are used in academic research by making it more difficult to reverse-search tweet text. Changing content of tweets for purposes besides formatting may violate the Twitter (2018) developer agreement. However, Fiesler and Proferes’ (2018) survey of Twitter users’ attitudes toward tweets being used in academic research made it clear that people strongly prefer that their tweets be anonymized. Furthermore, because I work in Europe and have collected tweets authored by Europeans whose data subject rights are protected by the General Data Protection Regulation, it is consistent within my institutional and national context to provide additional layers of anonymity to authors—even where they have consented via Twitter Terms of Service to have their content published. I maintain the original text of all tweets cited in this study and can share with researchers on request.

2. I follow Huddleston and Pullum’s (2002) use of the label “gerund-participle” rather than “gerund” or “present participle.” They argued that the distinction in traditional grammars between gerunds and present participles is not meaningful in Modern English (Huddleston & Pullum 2002:75-83; see also De Smet 2014).


4. A targeted search of Twitter shows that require can also be used in the AEP. A tweet from an unknown location provides the example, “Surely Lineker requires fired for that.” See also Duncan (2021).

5. Versions of the scripts I wrote to interface with twitteR are freely available for download at https://files.warwick.ac.uk/cstrelluf/browse/Big_Data. Since collecting the dataset for this study I have improved these scripts so that many of the characters and other formatting issues
in tweets that are problematic for analysis in R are screened out during sampling. This removes much of the need for aggressive cleansing of a corpus after the sample is created and results in more tweets being retained.

6. The editors rightly point out that *unlocked* is a canonical example of a participle form that is ambiguous as to whether it is verbal or adjectival (e.g., “They unlocked the door” versus “The door is unlocked”). I interpret these examples as verbal because, in discourse about English football, *unlock* is routinely used as an eventive verb to describe an action a team or manager should take to cause a player to realize their potential. In the context of this example, a Google search for “unlock Pogba” will reveal many examples of this usage.

7. This analysis is limited to interruption by -ly adverbs. There are a small number of interesting cases of interruptions by adverbial uses of *better, further,* and *well,* among others. These instances require more nuanced analysis than I am able to provide here, but as a group align with -ly adverbs in occurring more with *need+TO* than *need+ING* and *need+ED*. On the other hand, the intensifier *fucking,* as in (21), occurs more frequently with *need+ED* and *need+ING* than *need+TO*. This will be explored in future work.

**Software**


**References**


**Author Biography**

**Christopher Strelluf** is an associate professor in the Department of Applied Linguistics at the University of Warwick. His research interests include language variation and change, sociophonetics, dialectology, and innovative methods in linguistic research.