Regional variation and syntactic derivation of low-frequency *need*-passives on Twitter

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. 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. 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.

Regional variation and syntactic derivation of low-frequency *need*-passives on Twitter

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[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, Frazier & Simon 1996; Teney 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 other grammars 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 study, and find that *need*+ING in tweets 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.

Put together, these findings highlight the value of increased and expanded examination of low-frequency variables like *need*-passives in dialectological, sociolinguistic, and micro-syntactic, 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. I argue that linguistic theory and knowledge of language will be improved through greater attention to low-frequency features.

2. Background

This section provides context for the dialectogical, 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 indicate 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. Production data from Twitter 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.

　　Need is the most productive and acceptable member of a set of verbs which may take past participles as passive complements. Edelstein (2014) labels this construction the "alternative embedded passive" (AEP). It is well established that three verbs license the AEP: *need* (Murray, Frazer & Simon 1996), *want* (Murray & Simon 1999), and *like* (Murray & Simon 2002). Additional verbs that have been cited as taking past participle passive complements include *could use* (Teney 1998:596, en. 4), *could stand (*LinguistList listserv conversation among Doug Wilson, Laurence Horn, and Charles C.

Doyle archived at http://listserv.linguistlist.org/pipermail/ads-l/2012-July/120999.html), *love, hate*, and *deserve* (Duncan 2019:3, fn. 2).

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.

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. Teney (1998:596, en. 5) likewise argued that *need*+ED is derived by a different syntax from *need*+TO. Teney 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 (Teney 1998:596, en. 5).

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.[3] 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 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 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 p.c.). (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 derivation. Figures were created with LingTree by SIL International [2020].)

[FIGURE 1 HERE]

[FIGURE 2 HERE]

Edelstein's 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 Teney (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 adjectival. Edelstein argues (2014:265), "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 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) include *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 description (and treating *need*+ING as a raising construction like *need*+TO, so that its non-finite phrasal complement will be a tense phrase).

[FIGURE 3 HERE]

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) categorize *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, fn. 2). Generally it seems that the AEP and concealed passive can be formed from an identical set of matrix verbs.[4]

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 and 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]."

Doyle (2014:104), following the implication of complementarity in Murray, Frazer, and Simon (1996), 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" (2014:104-105), providing empirical evidence that *need*+ING and *need*+ED are complementary variants in the United States.

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 to those of *need*+ED (e.g., Teney 1998; Edelstein 2014), but also in small rhetorical moves, such as Murray, Frazer, and Simon's (1996:266) positioning of *need*+ED as a "regional" alternative to *need*+ING, and Labov, Ash, and Boberg's (2006:293) juxtaposition of

need+ED against "other dialects" that use *need*+ING or *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 (2014:232; see also 2013:85) mentions variation in written Englishes between *want*+TO and *want*+ING from the fourteenth century in a broader examination of the collapse of the gerund/participle distinction in English, which naturally positions *want*+ING as indicative of larger patterns in English.

2.3. Transitive-*need*

De Smet (2014:85-86) finds the earliest corpus attestations of both transitive-*need*+ED and transitive-*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) describe transitive-*need*+ED as "a raised object followed by an *-ed* participle clause," exemplified by the sentence reprinted here as (7). They give no indication that the construction is anything but standardly available across Englishes.

(7)   I want/need this watch repaired immediately. (example [16.54ii] from Quirk,
        Greenbaum, Leech & Svartvik 1985:1207)

Huddleston and Pullum (2002:1245) cite both transitive-*need*+ED and transitive-*need*+ING as the "concealed passive in a complex catenative construction" with the examples reprinted as (8)-(9).

(8)   He needs/wants his hair cut. (extracted from example [60iv] in Huddleston & Pullum
        2002:1245)

(9)   He needs/wants his hair cutting. (extracted from example [60iv] in Huddleston &
        Pullum 2002:1245)

However, in contrast to Quirk, Greenbaum, Leech, and Svartvik (1985), Huddleston and Pullum (2002:1245) describe transitive-*need*+ED in (8) "as an alternant of the concealed passive," which is "restricted to certain regional dialects such as Scottish."

2.4. *Need*-passives and Other Low-Frequency Features

        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, during sociolinguistic interviews I conducted in Kansas City (reported in Strelluf 2018), 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, p.c.). 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 (1966/2006:32) 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 2012:206-241) 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) note 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) liken *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: "Evidently, 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, though, 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 points 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 suggests 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, points 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 multi-dialectal 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., Corbett, Hippisley, Brown & Marriott 2001; Bybee & Thompson 1997) 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 "Conserving Effect" (1997) 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 corpora of productions 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, 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 5 July and 4 September 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 accordingly demonstrated that geographically associated tweets can generate robust dialect maps (e.g., Eisenstein, 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). Tagging procedures were detailed in Strelluf (2020), but

I repeat here that, as I developed methods for this project, taggers nearly always failed to tag *need*+ED--usually tagging passive complements as nouns. In this project, I prevented the tagger from coding *need* lexemes as verbs, and this 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 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 6,984 *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 (13-15), and other instances of text after *need* being tagged as a verb due to misspellings and unconventional formatting (16-17) 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 STRAAAAAAAAAAAAIT 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)-(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 renogatiating because they are part of deals. (Manchester; correctly tagged as *need*+ING)

A few tweets were correctly tagged as containing a *need*-passive, but 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)-(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 the 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 Teney 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 64 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

This section begins by comparing proportions of *need*-passives. That is followed by an examination of transitive-*need*. The section concludes with an exploration of syntactic characteristics of *need*-passives.

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.

[FIGURE 4 HERE]

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 the 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 principle components. In Figure 5, *need*+ING corresponds to the x-axis and *need*+ED the y-axis.

[FIGURE 5 HERE]

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 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:

(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 ($r=0.834$; $p<0.001$). 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 ($p=0.245$).

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 Teney (1998) and Edelstein (2014) to tweets. In the case of *need*+ED, these tests will check whether productions in Twitter align with Teney's and Edelstein's (primarily) judgment-based data. In the case of *need*+ING, the tests will check whether *need*+ING follows a similar set of syntactic constraints to *need*+ED.

The first two tests I apply were used by Teney (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.

Teney (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)

Teney (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)-(30).

(29)   Just like Pereira would unlock Pogba, now Lingard needs unlocked too? (Liverpool)

(30)   Looks like Klopp needs unlocking. (Liverpool)[6]

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 Teney's and Edelstein's claims that *need*+ED allows only verbal complements. In doing so, *need*-passive productions on Twitter offer no challenge to the conclusions Teney and Edelstein each reach from acceptability judgments about *need*+ED syntax. Indeed, the extension of Teney's and Edelstein's 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)-(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 above), 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 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). 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.[7]

(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)

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).

(37)   *The dogs need not walked. (example [62a] in Edelstein 2014:264)

(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 50 tweets.

Edelstein's 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 analysis of *need*+ED syntax.

It seems plausible that Edelstein's 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 superficially in participle morphology.

5. Discussion

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 (1958) 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. Furthermore on this view, 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 proposes for the AEP. Indeed, Huddleston and Pullum's description suggests that *need*+ING and transitive-*need*+ING (2002:1206) 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). 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.

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 adverbs exactly as they appear in tweets, I have changed about 20 percent of words elsewhere in tweets. This procedure is 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's (2018) survey of Twitter users' attitudes toward tweets being used in academic research makes 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 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 argue that the distinction in traditional grammars between gerunds and present participles is not meaningful in Modern English (2002:75-83; see also De Smet 2014).

3. See Adger (2003:315-326) for an accessible overview of raising and control constructions.

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."

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 players 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.

References

Adger, David. 2003. *Core syntax: A Minimalist approach*. Oxford: Oxford University Press.

Bloomquist, Jennifer. 2009. Dialect differences in central Pennsylvania: Regional dialect use and adaptation by African Americans in the lower Susquehanna Valley. *American Speech* 84(1). 27-47.

Boberg, Charles. 2010. *The English language in Canada: Status, history and comparative analysis*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781056

Bontcheva, Kalina, Leon Derczynski, Adam Funk, Mark A. Greenwood, Diana Maynard & Niraj Aswani. 2013. TwitIE: An open-source information extraction pipeline for microblog text. In Galia Angelova, Kalina Bontcheva & Ruslan Mitkov (eds.), *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, 83-90. Hissar, Bulgaria: RANLP. Available at https://gate.ac.uk/sale/ranlp2013/twitie/twitie-ranlp2013.pdf

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan & Paul Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In Joan Bybee & Paul Hopper (rds.), *Frequency and the emergence of linguistic structure*, 1-24. Amsterdam: John Benjamins.

Bybee, Joan L. & Sandra A. Thompson. 1997. Three frequency effects in syntax. *Berkeley Linguistics Society* 23. 65-85.

Chambers, J. K. 1995. The Canada–U.S. border as a vanishing isogloss: The evidence of chesterfield. *Journal of English Linguistics* 23. 155-166.

Corbett, Greville, Andrew Hippisley, Dunstan Brown & Paul Marriott. 2001. Frequency, regularity and the paradigm: A perspective from Russian on a complex relation. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 201-226. Amsterdam: John Benjamins.

Cox, Felicity & Sallyanne Palethorpe. 2007. Australian English. *Journal of the International Phonetic Association* 37(3). 341-350. doi:10.1017/S0025100307003192

De Smet, Hendrik. 2013. Change through recombination: Blending and analogy. *Language Sciences* 40. 80–94. http://dx.doi.org/10.1016/j.langsci.2013.03.003

De Smet, Hendrik. 2014. Constrained confusion: The gerund/participle distinction in Late Modern English. In Marianne Hundt (ed.), *Late Modern English syntax*, 224-238. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139507226.017

Derczynski, Leon, Diana Maynard, Niraj Aswani & Kalina Bontcheva. 2013. Microblog-genre noise and impact on semantic annotation accuracy. In Gerd Stumme & Andreas Hotho (eds.), *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, 21-30. New York: Association for Computing Machinery.

Derczynski, Leon, Alan Ritter, Sam Clark & Kalina Bontcheva. 2013. Twitter part-of-speech tagging for all: Overcoming sparse and noisy data. In Galia Angelova, Kalina Bontcheva & Ruslan Mitkov (eds.), *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, 198-206. Hissar, Bulgaria: RANLP. Available at http://www.derczynski.com/sheffield/papers/twitter_pos.pdf

Doyle, Gabriel. 2014. Mapping dialectal variation by querying social media. In Shuly Wintner, Sharon Goldwater & Stefan Riezler (eds.), *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 98-106. Gothenburg, Sweden: Association for Computational Linguistics. Available at http://aclweb.org/anthology/E/E14/E14-1011.pdf

Duncan, Daniel. 2019. Grammars compete late: Evidence from embedded passives. *University of Pennsylvania Working Papers in Linguistics* 25(1). 1-10. Available at https://repository.upenn.edu/pwpl/vol25/iss1/11/

Edelstein, Elspeth. 2014. This syntax needs studied. In Raffaella Zanuttini & Laurence Horn (eds.), *Micro-syntactic variation in North American English*, 242-268. Oxford: Oxford University Press.

Eisenstein, Jacob. 2017. Written dialect variation in online social media. In Charles Boberg, John Nerbonne & Dominic Watt (eds.), *Handbook of dialectology*, 368–383. Hoboken, NJ: Wiley.

Eisenstein, Jacob, Brendan O'Connor, Noah A. Smith & Eric P. Xing. 2012. Diffusion of lexical change in social media. *PLOS ONE* 9(11). e11 3114. doi:10.1371/journal.pone.0113114

Fiesler, Casey & Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media + Society* 4(1). 1-14. https://doi.org/10.1177/20563051187633

Gentry, Jeff. 2015. twitteR: R Based Twitter Client. R package version 1.1.9, retrieved Nov 2016 from https://CRAN.R-project.org/package=twitteR

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury & Peter Trudgill. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.

Grieve, Jack, Andrea Nini & Diansheng Guo. 2018. Mapping lexical innovation on American social media. *Journal of English Linguistics* 46(4). 293-319. doi:10.1177/0075424218793191

Hartigan, J.A. 1975. *Clustering algorithms*. New York: Wiley.

Hartigan, J. A. & M. A. Wong. 1979. A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28(1). 100-108. doi: 10.2307/2346830

Hickey, Raymond. 2007. *Irish English: History and present-day forms*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511551048

Hickey, Raymond. 2018. A survey of Irish English usage: Choropleth maps. Available at https://www.uni-due.de/IERC/

Holmes, Janet & Nick Wilson. 2017. *An introduction to sociolinguistics*. 5th edn. London: Routledge.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jamieson, E, Shouchun Chien, Gary Thoms, David Adger, Caroline Heycock & Jennifer Smith. 2019. When intuitions (don't) fail: Sociosyntax in the analysis of Scots. Paper presented at UKLVC12, Queen Mary University and University College-London, London, UK.

Jones, Taylor. 2015. Toward a description of African American Vernacular English dialect regions using "Black Twitter." *American Speech* 90(4). 403-440. doi:10.1215/00031283-3442117

Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In Randolph Quirk & Henry G. Widdowson (eds.), *English in the world: Teaching and learning the language and literatures*, 11- 30. Cambridge: Cambridge University Press.

Kachru, Braj B. 2005. *Asian Englishes: Beyond the canon*. Hong Kong: Hong Kong University Press.

Kachru, Braj B., Yamuna Kachru & Cecil L. Nelson. 2006. *The handbook of World Englishes*. Malden, MA: Blackwell.

Kaschak, Michael P. 2006. What this construction needs is generalized. *Memory & Cognition* 34(2). 368-379.

Kaschak, Michael P. & Arthur M. Glenberg. 2004. This construction needs learned. *Journal of Experimental Psychology: General* 133(3). 450-467.

Kassambara, Alboukadel & Fabian Mundt. 2020. factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7, retrieved March 2020 from https://CRAN.R-project.org/package=factoextra

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: De Gruyter.

Lim, Lisa. 2012. Standards of English in South-East Asia. In Raymod Hickey (ed.), *Standards of English: Codified varieties around the world*, 274-293. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139023832.015

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

Maher, Zach & Jim Wood. 2011. Needs washed. Yale Grammatical Diversity Project: English in North America. Available online at http://ygdp.yale.edu/phenomena/needs-washed. Accessed on 2019-12-03. Updated by Tom McCoy (2015) and Katie Martin (2018).

Methrie, Rajend. 1992. *English in language shift: The history, structure and sociolinguistics of South African Indian English*. Cambridge: Cambridge University Press.

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

Montgomery, Michael 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*, 189-212. Tuscaloosa: University of Alabama Press.

Murray, Thomas E., Timothy C. Frazer & Beth Lee Simon. 1996. *Need*+past participle in American English. *American Speech* 71(3). 255-271.

Murray, Thomas E. & Beth Lee Simon. 1999. *Want*+past participle in American English. *American Speech* 74(2). 140-164.

Murray, Thomas E. & Beth Lee Simon. 2002. At the intersection of regional and social dialects: The case of *like*+past participle in American English. *American Speech* 77(1). 32-68.

Murray, Thomas E. & Beth Lee Simon. 2006. What is dialect? Revisiting the Midland. InThomas E. Murray and Beth Lee Simon (eds.), *Language variation and change in the American Midland*, 1-30. Amsterdam: John Benjamins.

Pavalanathan, Umashanthi & Jacob Eisenstein. 2015. Audience-modulated variation in online social media. *American Speech* 90(2). 187-213. doi:10.1215/00031283-3130324.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R

    Foundation for Statistical Computing www.r-project.org/ (March 2020).

Rousseeuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster

    analysis. *Journal of Computational and Applied Mathematics* 20. 53-65. doi:10.1016/0377-

    0427(87)90125-7

SIL International. 2020. LingTree [Computer program]. Version 1.2.4, retrieved Jan 2020 from

    https://software.sil.org/lingtree/

Smith, Jennifer, David Adger, Brian Aitken, Caroline Heycock, E Jamieson & Gary Thoms. 2019. *The*

    Scots syntax atlas. Glasgow: University of Glasgow. Available at https://scotssyntaxatlas.ac.uk

Squires, Lauren (ed.). 2016a. *English in computer-mediated communication: Variation, representation,*

    and change [Topics in English Linguistics 93]. Berlin: De Gruyter.

Squires, Lauren. 2016b. Twitter: Design, discourse, and the implications of public text. In Alexandra

    Georgakapoulou & Tereza Spilioti (eds.), *Routledge handbook of language and digital*

    communication, 239-255. New York: Routledge.

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

Strelluf, Christopher. 2018. *Speaking from the Heartland: The Midland vowel system of Kansas City*

    [Publication of the American Dialect Society 103]. Durham, NC: Duke University Press.

Strelluf, Christopher. 2019. *anymore,* It's on Twitter: Positive-*anymore,* American regional dialects, and

    polarity-licensing in tweets. *American Speech* 94(3). 313-351.

Strelluf, Christopher. 2020. *needs*+PAST PARTICIPLE in US and UK regional Englishes on Twitter.

    World Englishes 39(1). 119-134. doi:10.1111/weng.12451

Tatman, Rachel. 2018. Should you keep the tweet?: Balancing reproducibility, open data and

    participant privacy. Workshop session presented at New Ways of Analyzing Variation [NWAV]

    47, University of Wisconsin, Madison, WI. Available at

    http://www.rctatman.com/files/Tatman_2018_NWAV_ShouldYouKeepTheTweet.pdf

Tenny, Carol. 1998. Psych verbs and verbal passives in Pittsburghese. *Linguistics* 36(3). 591-597.

Thorndike, Robert L. 1953. Who belongs in the family? *Psychometrika* 18(4). 267-276.

Tibshirani, Robert, Guenther Walther & Trevor Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B* 63(Part 2). 411-423.

Trudgill, Peter. 1983. On dialect: Social and geographical perspectives. Oxford: Blackwell.

Trudgill, Peter. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Oxford: Oxford University Press.

Twitter. 2018. Developer agreement. https://developer.twitter.com/en/developer-terms/agreement (1 June, 2018).

Youmans, Gilbert. 1986. Any more on *anymore*: Evidence from a Missouri regional dialect survey. *American Speech* 61(1). 61-75.

Appendix 1

Variety	City	Counts				Proportions		
		needs+TO	*needs*+ING	*needs*+ED	Total	TO	ING	ED
UK	Aberdeen	110	25	76	211	52.1	11.8	36.0
	Belfast	160	25	200	385	41.6	6.5	51.9
	Birmingham	589	461	6	1056	55.8	43.7	0.6
	Cardiff	432	257	6	695	62.2	37.0	0.9
	Edinburgh	485	140	396	1021	47.5	13.7	38.8
	Glasgow	777	200	674	1651	47.1	12.1	40.8
	Leeds	656	790	22	1468	44.7	53.8	1.5
	Liverpool	836	1137	25	1998	41.8	56.9	1.3
	London	1433	534	28	1995	71.8	26.8	1.4
	Manchester	1418	1584	28	3030	46.8	52.3	0.9
	Newcastle	529	360	210	1099	48.1	32.8	19.1
	Northampton	123	112	2	237	51.9	47.3	0.8
	Norwich	214	135	1	350	61.1	38.6	0.3
	Nottingham	595	572	14	1181	50.4	48.4	1.2
	Peterborough	101	74	2	177	57.1	41.8	1.1
	Plymouth	95	72	3	170	55.9	42.4	1.8
	Southampton	247	168	2	417	59.2	40.3	0.5
US	Atlanta	762	42	6	810	94.1	5.2	0.7
	Birmingham	538	53	19	610	88.2	8.7	3.1
	Boston	1796	136	26	1958	91.7	6.9	1.3
	Chicago	1435	108	26	1569	91.5	6.9	1.7
	Cleveland	568	35	110	713	79.7	4.9	15.4
	Columbus	554	43	174	771	71.9	5.6	22.6
	Dallas	1487	111	34	1632	91.1	6.8	2.1
	Denver	681	63	24	768	88.7	8.2	3.1
	Detroit	666	35	27	728	91.5	4.8	3.7
	Indianapolis	937	48	194	1179	79.5	4.1	16.5
	Kansas City	848	47	97	992	85.5	4.7	9.8
	Los Angeles	911	78	17	1006	90.6	7.8	1.7
	Minneapolis	1076	91	17	1184	90.9	7.7	1.4
	New York	969	58	5	1032	93.9	5.6	0.5
	Philadelphia	1797	206	34	2037	88.2	10.1	1.7
	Phoenix	1838	125	59	2022	90.9	6.2	2.9
	Pittsburgh	1118	87	463	1668	67.0	5.2	27.8
	San Francisco	1539	143	11	1693	90.9	8.4	0.6
	Seattle	2191	290	57	2538	86.3	11.4	2.2
	St Louis	1097	81	57	1235	88.8	6.6	4.6
world	Auckland	273	74	1	348	78.4	21.3	0.3
	Cape Town	471	32	1	504	93.5	6.3	0.2
	Dublin	1034	219	17	1270	81.4	17.2	1.3
	Georgetown	4	1	0	5	80	20	0
	Hong Kong	180	26	2	208	86.5	12.5	1.0
	Islamabad	387	23	0	410	94.4	5.6	0.0
	Lagos	576	81	2	659	87.4	12.3	0.3
	Manila	455	115	4	574	79.3	20.0	0.7
	New Delhi	1640	77	5	1722	95.2	4.5	0.3
	Singapore	510	76	6	592	86.1	12.8	1.0
	Sydney	1494	286	11	1791	83.4	16.0	0.6

| Toronto | 1894 | 223 | 21 | 2138 | 88.6 | 10.4 | 1.0 |
| Vancouver | 1142 | 176 | 10 | 1328 | 86.0 | 13.3 | 0.8 |

Appendix 2

Variety	City	Counts			Proportions	
		Transitive-need+ED	Transitive-need+ING	Total	Transitive-need+ED	Transitive-need+ING
UK	Belfast	2	1	3	66.7	33.3
	Birmingham UK	3	3	6	50	50
	Cardiff	2	5	7	28.6	71.4
	Edinburgh	6	1	7	85.7	14.3
	Glasgow	12	1	13	92.3	7.7
	Leeds	0	1	1	0	100
	Liverpool	2	6	8	25	75
	London	4	3	7	57.1	42.9
	Manchester	2	13	15	13.3	86.7
	Newcastle	7	1	8	87.5	12.5
	Northampton	0	4	4	0	100
	Norwich	1	1	2	50	50
	Nottingham	4	5	9	44.4	55.6
	Peterborough	1	3	4	25	75
	Plymouth	0	1	1	0	100
	Southampton	1	1	2	50	50
US	Atlanta	11	0	11	100	0
	Birmingham US	8	0	8	100	0
	Boston	16	0	16	100	0
	Chicago	20	1	21	95.2	4.8
	Cleveland	10	0	10	100	0
	Columbus	9	0	9	100	0
	Dallas	24	1	25	96	4
	Denver	3	0	3	100	0
	Detroit	9	0	9	100	0
	Indianapolis	14	0	14	100	0
	Kansas City	10	2	12	83.3	16.7
	Los Angeles	3	0	3	100	0
	Minneapolis	8	0	8	100	0
	New York	8	0	8	100	0
	Philadelphia	24	3	27	88.9	11.1
	Phoenix	16	0	16	100	0
	Pittsburgh	9	0	9	100	0
	San Francisco	11	1	12	91.7	8.3
	Seattle	20	0	20	100	0
	St Louis	15	0	15	100	0
world	Auckland	2	0	2	100	0
	Cape Town	4	0	4	100	0
	Dublin	6	1	7	85.7	14.3
	Hong Kong	0	1	1	0	100
	Lagos	3	0	3	100	0
	Singapore	3	1	4	75	25
	Sydney	6	0	6	100	0
	Toronto	22	2	24	91.7	8.3
	Vancouver	16	1	17	94.1	5.9

Figure 1: Derivational structure of *need*+ED recreated from Edelstein (2014:265) for "The cat needs fed."
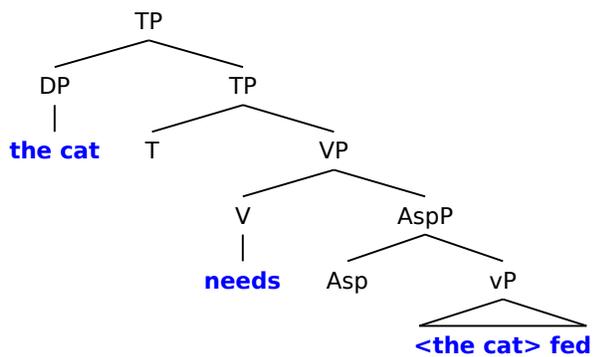


```
                    TP
             ┌──────┴──────┐
            DP            TP
             │      ┌──────┴──────┐
          the cat   T            VP
                           ┌──────┴──────┐
                           V            AspP
                           │      ┌──────┴──────┐
                         needs   Asp           vP
                                          ┌────┴────┐
                                       <the cat> fed
```

Figure 2: Derivational structure of *need*+TO recreated from Adger (2003:318) for "The cat needs to be fed"
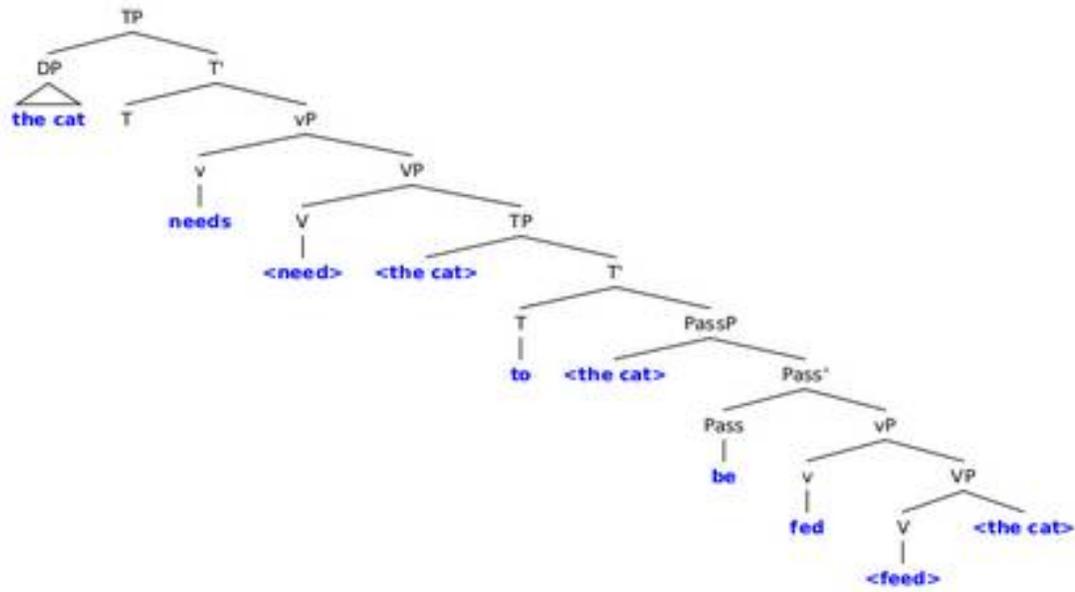
Figure 3: Derivational structure of *need*+ING based on description in Huddleston and Pullum
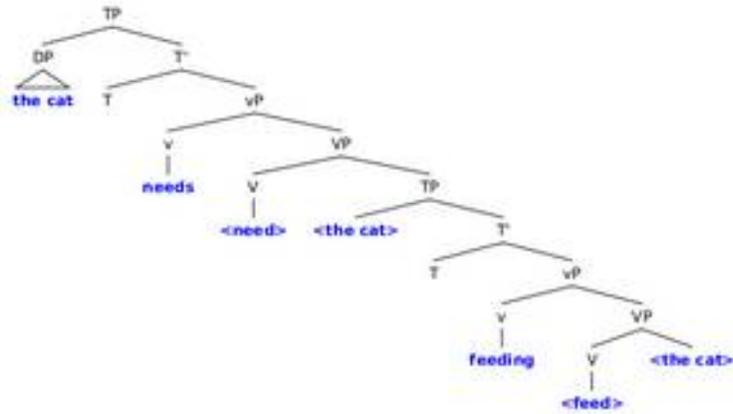
(2002:1999-1200) for "The cat needs feeding"

Figure 4: Proportions of *need*-passives in all cities, sorted by strength of preference for *need*+ED or
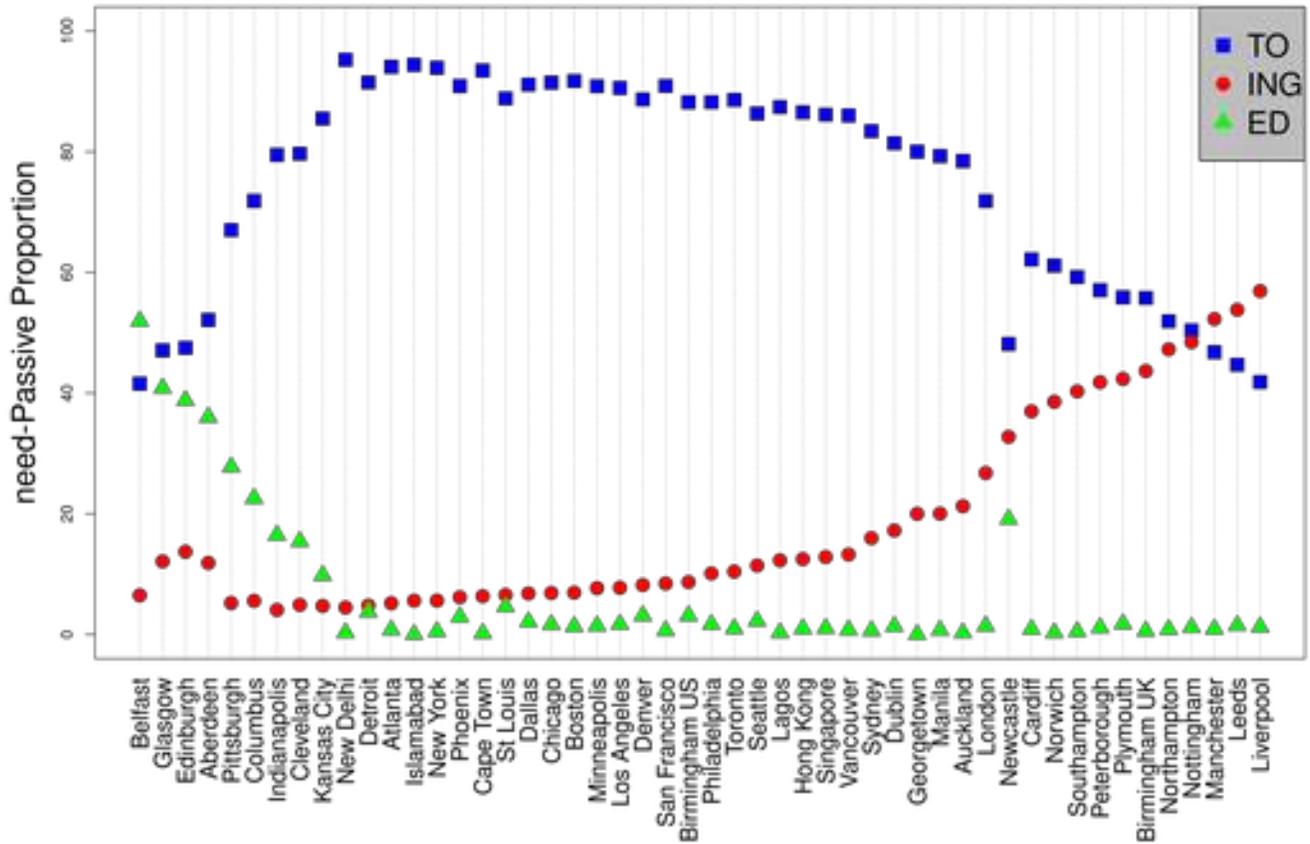
need+ING

Figure 5: Cluster analysis of *need* -passives in all cities