CrossMark

Language in Society **44**, 705–731.
doi:10.1017/S0047404515000603

Quantifying the referential function of general extenders in North American English

SUZANNE EVANS WAGNER

Department of Linguistics and Languages, Michigan State University, East Lansing, MI 48824, USA
wagnersu@msu.edu

ASHLEY HESSON

Department of Linguistics and Languages, Michigan State University, East Lansing, MI 48824, USA
bartell6@msu.edu

KALI BYBEL

Department of Linguistics and Languages, Michigan State University, East Lansing, MI 48824, USA
bybelkal@msu.edu

HEIDI LITTLE

Department of Linguistics and Languages, Michigan State University, East Lansing, MI 48824, USA
littlehe@msu.edu

ABSTRACT

Discourse markers (*like, I don't know*, etc.) are known to vary in frequency across English dialects and speech settings. It is difficult to make meaningful generalizations over these differences, since quantitative discourse-pragmatic variation studies 'lack [a] coherent set of methodological principles' (Pichler 2010:582). This has often constrained quantitative studies to focus on the form, rather than the function of discourse-pragmatic features. The current article employs a novel method for rigorously identifying and quantifying the referential function (set-extension) of general extenders (GEs), for example, *and stuff like that, or whatever*. We apply this method to GEs extracted from three corpora of contemporary North American English speech. The results demonstrate that, across varieties, (i) referential GEs occur at a comparable proportional rate in vernacular speech, and (ii) referential GEs are longer than nonreferential GEs. Collectively, these findings represent a step towards comparative quantitative studies of GEs' functions in

© Cambridge University Press, 2015 0047-4045/14 $15.00

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

discourse. (Discourse-pragmatic variation, general extenders, methodological approaches, American English, Canadian English)

INTRODUCTION

Discourse-pragmatic features such as fillers (e.g. *um, uh*), tag questions (e.g. *you know what I mean?*) and discourse particles (e.g. *like, well, so*) have been of great interest to variationist sociolinguists (e.g. Tagliamonte 2005; D'Arcy 2006; Cheshire, Kerswill, Fox, & Torgersen 2011 inter alia), but in general the study of discourse-pragmatic variation has long been a mostly qualitative enterprise (e.g. Schiffrin 1994). This is due to the many methodological obstacles that quantitative researchers face when attempting to circumscribe discourse-pragmatic features for their purposes. Importantly, quantitative sociolinguists working in the variationist paradigm (Labov 1966/2006; Tagliamonte 2012) rely on replicable definitions of sociolinguistic variables—or 'two or more ways of saying the same thing'—that can be applied to multiple datasets. Such definitions can be difficult to generate for discourse-pragmatic variables (Cheshire 1987; Dines 1980; Lavandera 1978). One of two approaches may be adequate, however: (i) counting all instances of a particular lexical FORM (e.g. *well, you know*) without regard to its function(s) in discourse, or (ii) counting all of the lexical forms that perform a single FUNCTION (e.g. verbs such as *say, be like, go* that perform a quotative function; see Walker (2010) for an overview of form vs. function-based approaches). Furthermore, circumscribing the envelope of variation is especially problematic for GEs given the difficulty of identifying where they could have occurred but did not.

Previously conducted analyses have largely been limited to the first approach, which identifies FORMS rather than their FUNCTIONS. Indeed, as Pichler (2010:596–97) points out, 'despite its great hermeneutic and explanatory values, consideration of function is not an integral design feature in all discourse variation studies', most likely because coding and quantifying discourse variants by function is a daunting task for many variables. A given discourse particle can functionally contribute to interaction in numerous ways, ranging from the negotiation of disfluency to interpersonal face-work (Fischer 1998). In dealing with this multifunctionality, the researcher must either (a) subjectively assign a 'primary' function to tokens that are demonstrably multifunctional, or (b) attempt to code all of a given token's pragmatic functions——an exercise that, like (a), may introduce unacceptable levels of subjectivity. Nonetheless, these problems must be surmounted if variationist researchers are to reap the benefits of an understanding of pragmatic function. For example, if individual functions can be isolated and tracked with respect to a discourse-pragmatic variable, such patterns can inform researchers' hypotheses on the variable's diachronic evolution. Synchronically, researchers can reliably observe whether a particular discourse function is instantiated using similar or different lexical forms across dialects, as has been successfully achieved for verbs that

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

perform a quotative function (*be like, say, go*, etc.) across global Englishes (e.g. Buchstaller 2008; D'Arcy 2012; though see Singler 2001 for a criticism of quotative analysis). Not all discourse-pragmatic variables, however, lend themselves as readily to a circumscription of function as do quotative verbs. The present work represents a preliminary effort to take a function-based approach (as in (ii) above) to the analysis of a more intractable variable: the general extender.

General extenders (henceforth GEs) in English include such utterance-final phrases as *and stuff like that, or whatever*, and *and all that kind of thing*. As illustrated in (1), they prototypically refer to items (*panettone, torrone*) that could evoke and extend an inferable set ('Italian baked goods'). Henceforward we call this the 'referential' or 'extending' function of GEs, and we indicate referents using underlining in our examples.

(1) I guess there's no cakes or cookies but there's, you know, you need the panettone, you need the torrone, you need *stuff like that.* (LCS I044, Lucia)[1]

GEs, however, can also perform nonreferential interpersonal functions such as closing social distance between interlocutors (see e.g. Aijmer 1985; Overstreet & Yule 1997). This particular subset of functions has been identified through careful qualitative analysis of select examples, a process that is difficult to replicate among analysts and/or generalize between instances. In this article, we attempt to isolate just the referential function and to account for the GE forms that typically perform this function. As illustrated in (2), isolating referential GEs is not always straightforward, since the speaker's intended set or referents may be ambiguous to the analyst.

(2) some of the facts are misconstrued *or whatever*

(Fisher, 03656, 07398)

In (2), it is unclear whether 'misconstrued' is the constituent being referred to by *or whatever,* as a member of a set of potential ways to ambiguate facts (e.g. misunderstand, misconstrue, misinterpret). *Or whatever* might alternatively extend from 'facts' to a set of things that can be misconstrued (e.g. facts, meanings, messages), or from 'some of the facts are misconstrued' to a set of events relevant in this context: maybe events in which confusion occurs (e.g. some of the facts are misconstrued, some unfounded judgments are made).

We employ a novel method for identifying the subset of referential GE variants: a method that places value on replicability. We use three corpora from two regional varieties of spoken North American English to demonstrate how this can be achieved, while simultaneously beginning to explore function-based patterns in GE use across varieties. We perform logistic regression analyses to determine the lexical, syntactic, and discourse predictors of referential GEs. The strikingly similar results across all three corpora demonstrate the potential efficacy of our

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

circumscription method for quantitative variationists. We find that referential GEs share a tendency to be syntagmatically long (e.g. *or something like that*) rather than syntagmatically short (e.g. *or something*) across varieties of English. In our DISCUSSION section, we review the implications this finding might have for a theory of GE grammaticalization and for a more general understanding of the relationship between GE form and function.

GENERAL EXTENDERS

General extenders (GEs) form a class of discourse phenomena that in variationist sociolinguistic studies of English, as mentioned above, are identified by their distinctive lexical form. Prototypical GEs follow 'a basic template… where a connector is [typically] required, a quantifier and/or a generic is necessary, and the comparative is optional' (Tagliamonte & Denis 2010:336). The first component is a connector (*and*, *or*, or null), and it is usually followed by a quantifier (e.g. *every, some, any, no*) and/or a generic noun (e.g. *things, stuff, crap*). A comparative such as *like that, that kind of, of that sort* may precede or follow the generic. Thus, examples of GEs include *and stuff, or whatever, and all that crap*, and *stuff like that.* GEs occur principally in spoken discourse and are typically clause-final.

In this article we examine GEs in US English and Canadian English, but GEs occur in all varieties of English studied to date, including British (Cheshire 2007; Pichler & Levey 2010, 2011; Denis 2011; Martinez 2011), Irish (O'Keeffe 2004), Australian (Norrby & Winter 2002), New Zealand (Stubbe & Holmes 1995), and Trinidadian Creole (Youssef 1993). GEs in US English have been examined in Overstreet (1999) and Overstreet & Yule (1997) and in Canadian English by Tagliamonte & Denis (2010). For a comprehensive discussion of both qualitative and quantitative studies of GEs, see Cheshire (2007). Preferences for certain GE lexical forms (regardless of function, which we discuss below) vary synchronically across world Englishes. For example, the GE form *and that* is very common in British dialects (Cheshire 2007; Pichler & Levey 2011) but not in North American dialects. Tagliamonte & Denis (2010) find increasing replacement in apparent time of *thing* GEs by *stuff* GEs in Toronto, Canada; this trend has been taking off more slowly in Britain (Denis 2011). Despite this lexical variety, however, GEs in global Englishes consistently adhere to the basic combinatorial 'template' described above.

Identifying GE pragmatic functions

As shown in (1), GEs may require the listener to extend a set of previously mentioned referents, relying on their shared knowledge with the speaker to do so. General extenders are also 'general' insofar as they may perform other functions in the discourse and have a wide range of pragmatic interpretations (Ball & Ariel 1978; Aijmer 1985; Ward & Birner 1993; Channell 1994; Overstreet & Yule

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

1997 inter alia). They may perform textual functions such as introducing a new topic, signalling turn transitions, or focusing a discourse entity (Aijmer 2002); and they may also perform interpersonal functions such as indicating politeness or closing social distance (Aijmer 1985; Overstreet & Yule 1997). Some GE tokens may serve no identifiable purpose whatsoever but simply 'punctuate or bracket units of discourse' (Pichler & Levey 2011:453).

For the purposes of variationist quantitative analysis, an examination of GE function would require the researcher to consistently identify functions in a way that is replicable both intra-study (for internal consistency) and inter-study. This consistency requirement is of course not a necessary prerequisite for more qualitatively oriented work. We limit discussion in this article to the challenge of coding GE pragmatic function for quantitative variationist work.

As mentioned above, however, this is a very challenging enterprise. In the absence of other guidelines, researchers may be forced to rely on intuition to ascertain which function(s) a given GE is performing. In two different studies of GEs in British English, Cheshire (2007) and Martinez (2011) both found the intuition method to be unreliable. Cheshire (2007:183–85) noted that it was often impossible to assign a single function to multifunctional GE tokens, and that any identification of functions was open to a conflicting interpretation. Martinez (2011:2465–66) further pointed out that even in the best case, careful subjective analysis was difficult to perform for the large number of GE tokens necessary for a good quantitative analysis.

Another approach, subsequently adopted by Cheshire (2007) and by Tagliamonte & Denis (2010), is to constrain the task from (a) identifying any and all GE functions to (b) a binary identification scheme in which GEs performing entirely nonreferential functions are contrasted with all others. Pichler & Levey (2011:452) call this kind of GE, exemplified in (2), a 'punctor devoid of referential and pragmatic meanings'.[2] For (3), Tagliamonte & Denis (2010:339) note that the speaker [03] uses *and stuff* despite the topic of conversation being only the video and nothing else that he and his friends did for the teacher.

(3) [03] 'Cause we made like a video.
[1] Yeah, I saw it.
[03] You remember, right?
[1] Yeah.
[03] So ah, we made her a video *and stuff*.

(2c/M/16 & interviewer)
Tagliamonte & Denis (2010:338, ex. 16)

In order to identify 'punctors', the researcher must decide whether a given GE token has a referent. 'Referent' means an entity (or entities) in the preceding discourse that forms part of the GE's inferable set. For example, in (1), *panettone, torrone, stuff like that*, there are two explicit referents preceding the GE, and

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

these evoke the set of baked goods that are then extended by *stuff like that*. To identify punctors, the researcher isolates GE tokens for which no referent/category can be located in the discourse. Doing so allows, ideally, for a straightforward proportional analysis of the ratio of punctors to nonpunctors. Unfortunately, locating referents in discourse is by no means a straightforward matter. Indeed, Tagliamonte & Denis (2010:354–55) made a ternary distinction in their coding system between GEs with discernible referents, those with no discernible referents and 'tokens where no set is immediately apparent but, given the appropriate context, a set is potentially available' (Tagliamonte & Denis 2010:365, n. 12). But in order to decide whether a set is 'potentially available' to the listener, the researcher must make assumptions about the interlocutors' shared knowledge and/or the speaker's intentions, neither of which are accessible. For example, in (4), two American high schoolers discuss the senior yearbook.

(4) Author: [I]t must be tough on people buying it then cause they don't know what the
product's gonna look like. They have to trust you that-
M: Yeah, but I mean most people they just want the year- It's the idea that it is a yearbook-
L: It's a senior thing
M: You know, and it is a big senior thing, you get your name on it *and stuff*.

(LCS I007, Melissa and Lucia)

Conceivably the intended set could be 'getting your name on the senior yearbook, and other kinds of yearbook personalizations, such as getting your photo printed inside'. The set might be 'the set of all things on which seniors get their name printed around graduation, such as yearbooks, t-shirts, graduation invitations, and photos'. The set might also be 'big senior things [activities/concepts], such as getting your name on the yearbook or going to senior prom'. Under another interpretation, *and stuff* does not have a set-extending function here at all, but is an utterance-final, nonreferential tag intended to close social distance with the interviewer by indicating that all parties in the conversation know what seniors value. It might, alternatively, have a textual function in signalling the end of a turn, or have no pragmatic function at all. Without knowing the speaker's actual intentions, or the listener's existing knowledge/assumptions, a variety of interpretations are possible.

Without a clear and explicit definition of when a GE can be said to lack a referent—that is, a definition that does not rely upon researcher intuition—identification of punctors is open to wide interpretation, as are the results that rely on it. And indeed, the results are inconsistent across studies. Tagliamonte & Denis (2010:356) report that for Toronto English, 'GEs that have no set-marking function are rare'; Cheshire (2007:177), by contrast, reports rates as high as

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

56.6% non-set-marking for the GE *and that* in her Milton Keynes data.[3] Though these results may represent dialect differences, they may also point to inter-analyst subjective differences.

This subjectivity is justifiable if the goal is to explore and describe GEs' multiple referent types and pragmatic functions (e.g. Overstreet 1999). Yet quantitative analyses of referential vs. nonreferential GE distribution must be founded on a consistent definition of co-reference, and this is currently lacking. In the next section, we suggest a novel method of identifying referential GEs that we believe to be less dependent on subjective interpretation, and thus well-suited to quantitative, cross-corpora analysis.

NEW DIAGNOSTICS FOR REFERENTIAL GES

Like Cheshire (2007) and Martinez (2011), we found that multiple interpretations of a single GE token were possible. In response to these practical problems, we developed a scheme that follows Cheshire (2007) and Tagliamonte & Denis (2010:337) in contrasting referential (set-extending) and nonreferential GE tokens. However, whereas these authors isolated tokens for which no referent could be identified, we take the opposite approach of isolating tokens for which a referent CAN unambiguously be located. As we noted earlier, it is relatively easy to justify almost any preceding discourse entity as a potential referent, meaning that this approach most likely underestimates the number of referent-less GE tokens in the data. Below, we outline a coding scheme that possibly overestimates the number of referent-less GEs, but that allows for a high degree of certainty with respect to the GEs that DO have a referent. The proportion of maximally unambiguous referential GEs can then be calculated. We suggest that this conservative approach is more useful to quantitative variationists attempting to examine GEs across studies. In what follows, we refer to set-extending GEs as eGEs and non-set-extending GEs as nGEs.

eGEs are defined as GE tokens that perform, at minimum, a set-extending function in the discourse. We acknowledge that eGEs may also be performing other functions, but do not attempt to account for these in our scheme, leaving this for future work. Likewise, we do not attempt to identify or categorize the various nonreferential functions of nGEs, recognizing—from the difficulties experienced by other researchers—that these are beyond the scope of our scheme at present. We return to this issue later in the article.

We have organized our eGE-focal coding scheme into a binary branching tree. The tree (Figure 1) represents a hierarchically ordered progression of decisions. It relies upon the operationalization of three key concepts: REFERENT, LOCAL DOMAIN, and DISCOURSE CONTEXT. The scheme was developed by a six-member trained coding team (including the authors), all of whom had a background in the GE literature and some experience in identifying GE features. In the remainder of this section, we describe the coding process step-by-step with illustrative examples.

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

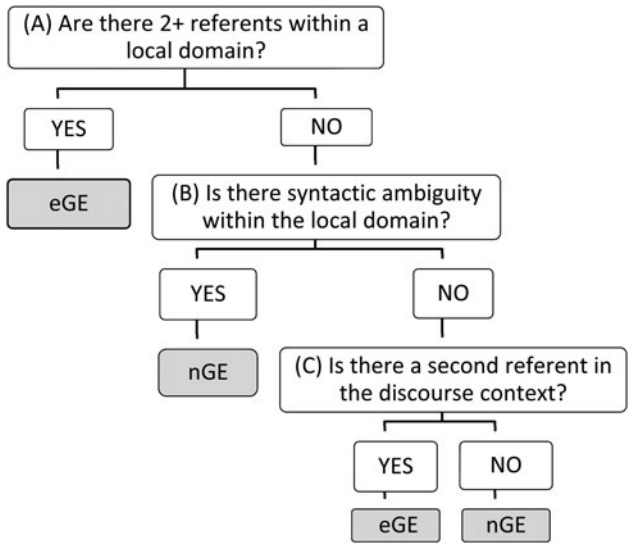


FIGURE 1. GE pragmatic function coding scheme.

Throughout, we employ the terminology currently used in generative syntax, derived from Chomksy's Minimalist program (Chomsky 1995).

Trained coders were directed to determine whether a given GE token had two or more identifiable REFERENTS (see question A) of any grammatical category (see ex. (5)), under the assumption that listing is good evidence for set-extension (Jefferson 1990). The speaker must show evidence of listing behavior in order to pragmatically convey that the list is incomplete or otherwise extendable. We follow Ward & Birner (1993) in defining a referent as a representative member of an inferable set to which a GE simultaneously refers. For example, in the phrase *parents or whatever,* parents serves as a representative member of some inferable set, possibly family members or 'people that you know', to which the GE *or whatever* refers.[4]

(5) and it could be like cousins brothers or sisters and parents *or whatever*.
(Fisher, 02552, 94267)

In every case of two or more referents, the GE was coded as extending (eGE), unless the referents could be considered synonyms (e.g. *fat and triglycerides and stuff*) or components of an idiom, such as *raining cats and dogs or whatever*.

In order to evaluate the number of referents a GE has, one must first be able to unambiguously identify what those referents are, a task that our coders (along with previous analysts) found difficult. In (6), a coder could either identify a VP (verb

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

phrase) referent do stuff with my grandfather or a DP (determiner phrase) referent my grandfather to which *or something* refers.

(6) She has to do stuff [$_{PP}$ with my grandfather *or something*].
(LCS I044, Melissa)

In order to resolve issues like this, we defined a standardized starting point for identifying referents: a LOCAL DOMAIN. We specified local domain as the prepositional phrase (PP) if one were present (see ex. (7)), or the tensed phrase (TP) in all other cases (see ex. (8)).

(7) I learnt stuff like that as a communications major all different studies [$_{PP}$ about the different ways people communicate and different depths and demographics *and stuff*]. (Fisher, 01894, 84404)
(8) [$_{TP}$ you have to say hello and go downstairs and give him a hug, *and all this*,] (TOR, I, §)

If only one potential referent occurred within the local domain, coders were directed to consider whether it was unambiguously the intended referent of the GE (9), or whether competing referents might be present (10) (see question B). The intended referent of the GE in (10) is not unambiguously identifiable as *and stuff* could potentially refer to the DP its phone name or it could refer to the VP buying its phone name. Competing referent cases were largely confined to those in TP domains with potential referents in different lexical categories, as seen in (10). Had we restricted local domains to just PPs, however, we would have not have been able to include many unambiguous eGE tokens in which two referents from the same lexical category (e.g. two VPs, two DPs) were clearly present.

(9) they just everything was just the same because I guess the hotel has its own security with cameras in the elevators and [$_{PP}$ in the hallway *and stuff*]
(Fisher, 08610, 66661)
(10) [$_{TP}$ But then, they're buying its phone name *and stuff,* so.] (TOR, 2, n)

Not only can we not be sure, in (10), which set the speaker intended to extend, it's also unclear that the speaker intended to extend a set at all. From this perspective, these GEs are likely to serve nonreferential functions, since interlocutors would have faced similar indeterminacy. Thus we classified syntactically ambiguous tokens as nGEs. If the single referent was deemed unambiguous, with no competitors in the local domain, as in (9), the DISCOURSE CONTEXT was used as a final potential source for a second referent to satisfy the 2+ referent requirement for unambiguously exhibiting listing behavior (see question C). We defined discourse context as the speaker turn—however determined by the transcription—and prevented coders from searching elsewhere in the discourse. The turn-level restriction

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

is somewhat stipulative, but necessary for reliability purposes. Without a defined unit of context, coders varied substantially on their willingness to include GE tokens with one or more clause-external referents. In example (9), repeated with discourse context as (11) below, there is an additional referent within the turn.

(11) Well nothing changed with security wise there. They just everything was just the same because I guess the hotel has its own security with cameras in the elevators and in the hallway *and stuff.* (Fisher, 08610, 66661)

GEs that either satisfied question (A) or passed the required diagnostics in questions (B) and (C) were coded as set-extending eGEs. Those that did not were coded as non-set-extending nGEs.

DATA

Data were drawn from three comparable corpora (Table 1). Our primary analysis is of the two US English corpora (LCS and Fisher), described below. Additional comparison data were drawn from a small sample of Canadian English from the Toronto English Archive.

Language change and stabilization (LCS)

The LCS corpus comprises sociolinguistic interviews with white females in 2005 and in 2006 (Wagner 2008). The participants were high school students aged sixteen to eighteen in 2005, and were either college freshmen or still in high school in 2006. They were long-time residents of Philadelphia, Pennsylvania,

TABLE 1. *Characteristics of the three corpora samples.*

	LCS	Fisher	TOR
Regional dialect	US English	US English	Canadian English
Speaker origin	Philadelphia, Pennsylvania, USA	Pennsylvania, USA	Toronto, Ontario, Canada
Year recorded	2005–2006	2003	2003–2010
N speakers	13 (recorded twice)	56	9
Speaker age	16–20	16–24	16–24
Typical interview length	1 hour	5–10 minutes	1 hour
Interlocutor	familiar fieldworker	stranger	familiar fieldworker
Format	sociolinguistic interview	telephone conversation	sociolinguistic interview
N GE tokens	605	337	272
N per 10,000 words	25.42	42.02	28.56

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

and native speakers of the local dialect of English. Thirteen speakers were selected on the availability of transcripts at the time of analysis.

Fisher

The Fisher corpus comprises short telephone conversations (five to ten minutes) between pairs of strangers from across the USA on a prescribed topic. The Fisher English corpus was compiled by the Linguistic Data Consortium in 2003 (Cieri et al. 2004, 2005). We selected a subset of participants whose demographic characteristics most closely matched the LCS speakers: fifty-six female speakers aged sixteen to twenty-four from Pennsylvania.

Toronto English Archive (TOR)

This archive contains sociolinguistic interviews conducted as part of several University of Toronto research projects (Tagliamonte 2003–2006, 2007–2010). The analyses in Tagliamonte & Denis (2010) are based on the entire corpus of eighty-seven speakers. Our subset comprises the participants whose demographic characteristics most closely matched the LCS speakers: nine female speakers aged sixteen to twenty-four.

RESULTS

Coding reliability

Prior to discussing the results gained from applying our coding scheme to the corpora introduced above, we first detail the results of various reliability tests conducted in order to confirm that our scheme accomplished its intended objective: operationalizing a consistent method by which to identify eGEs.

Scheme-based coding

We conducted two parallel coding runs on the Language Change and Stabilization (LCS) corpus of sociolinguistic interviews with teenage girls from Philadelphia (Wagner 2008). The same six-member coding team conducted each run on an individually assigned subset of data. Six subsamples were redistributed between passes such that individual coders did not code the same token twice. We did not alter the coding scheme between experimental runs. These runs constitute our SCHEME-BASED coding.

Intuition-based coding

We also recruited a group of twelve untrained coders, for the purpose of comparing their coding with our own. These coders performed what we call INTUITION-BASED CODING on tokens that had been extracted from the LCS by the scheme-based

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

team. They were asked to use their subjective intuition to code these tokens for their pragmatic function: set-extending or non-set-extending. These coders received a basic definition of the GE variable, focusing on the concept of set extension. They were not exposed to our scheme or any previous GE studies. Each of the twelve untrained individuals performed one pass on a subset of the data. Our six subsamples were also redundantly assigned to two untrained volunteers, resulting in two full passes of the full dataset.

Inter-rater reliability tests

We used three methods to assess inter-rater reliability. PERCENT MATCH (i) and COHEN'S KAPPA (ii) were used for both the scheme-based and intuition-based coding experiments. BOOTSTRAPPING (iii) was only applied to the scheme-based coding sample as a test of intra-rater reliability. For all three methods, a match is defined as two coders (either trained or untrained) agreeing on the pragmatic function of a GE, that is, both identifying it as an eGE or an nGE. A mismatch refers to disagreement between coders.

(i) PERCENT MATCH. For each experiment, we divided the total number of matches by the number of tokens in the sample (N = 605 for both), then multiplied the result by 100. This technique provided a rough estimate of agreement, but did so without factoring in the expected number of matches or the hypothetical distribution of eGEs/nGEs in our data.

(ii) COHEN'S KAPPA. We tabulated the number of nGE matches, eGE matches, nGE-eGE mismatches, and eGE-nGE mismatches between raters for both experiments. We subsequently calculated the Kappa value (0–1, where 1 is the highest possible agreement) for scheme-based coding and intuition-based coding respectively. This technique provided a measure of agreement that was sensitive to the expected number of matches by experiment, where the expected value was based on the total number of eGE and nGE responses. However, it did not account for coder-specific variation.

(iii) BOOTSTRAPPING. Bootstrapping is a statistical technique that creates a large, simulated distribution from random sampling of a smaller distribution (see Hinneburg, Mannila, Kaislaniemi, Nevalainen, & Raumolin-Brunberg 2007). As used here, it provides a basis for comparing each coder to chance, where 'chance' is represented by iterative samplings of possible outcomes from the coders' sample of responses. We used bootstrapping to ensure that individual coder's coding performance did not obscure the overall measures of reliability reported for our coding scheme.

Based on percent match alone, our scheme offers an increase in agreement over intuitive coding (Table 2). Scheme-based coders agreed on 79.7% of the tokens while intuition-based coders agreed on 61.8% of tokens. These percentages can be somewhat misleading, however, as they do not consider the binary nature of our pragmatic function distinction. With only two options (eGE and nGE),

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

chance alone may account for the majority of agreement observed in both cases. Cohen's Kappa accounts for some of this baseline agreement and allows us to compare the magnitude of agreement beyond that which is expected in the sample (Gwet 2012). It ranges from 0 to 1, where 1 represents perfect agreement between coders. Using this measure, the difference between intuition-based coding and scheme-based coding for pragmatic function remains modest. Our scheme-based coders agree with a Kappa (K) value of .337. Our intuition-based coders, by contrast, achieved an agreement of K = .243. These values both fall within the 'fair agreement' category of the traditional Kappa interpretation scale (Landis & Koch 1977), but represent opposite ends of the grouping (0–.20 = slight agreement, .21–.40 = fair agreement, and .41–.60 = moderate agreement). Cohen's Kappa contextualizes the difference between scheme- and intuition-based coding by highlighting the contribution of random agreement.

Our bootstrapping technique (Table 3) demonstrated that all of our scheme-based coders achieved a higher rate of agreement than would have been expected by pseudo-random assignment of eGE/nGE codes within our sample. In other words, their observed proportion of agreement was significantly higher than the average proportion of agreement in simulated distribution of coding runs. Though this is not particularly remarkable as an overall performance metric, since we expect the scheme-based coders to systematically agree, it shows that our scheme's reliability is not being unduly affected by a single coder or coding pair. Such consistency is not immediately apparent based on the individual coders' percent matches. Coder C, for instance, had an especially low percent match rate of 64.5%; this coder also logged fewer training hours than the other five coders. While the variability in percent match for individual coders indicates

TABLE 2. *Percent match and Cohen's Kappa for scheme-based and intuition-based coding.*

Experiment	Number of tokens (N)	Percent match (%)	Cohen's Kappa (K)
Scheme-based	605	79.7	0.337
Intuition-based	605	61.8	0.243

TABLE 3. *Bootstrapping results by members of the scheme-based coding team.*

Coder	Number of tokens (N)	Percent match (%)	*p*-value
A	78	87.2	< 0.001
B	83	80.1	< 0.001
C	110	64.5	< 0.001
D	129	87.6	< 0.001
E	81	86.4	< 0.001
F	124	75.0	0.001

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

a range of skill and training in applying the coding scheme, these individual differences are not so extreme as to put any single rater below chance agreement.

Overall, in the absence of an accepted standard for reliability in sociolinguistics, we take the convergence of our three reliability tests as sufficient evidence that the coding scheme presented here represents a substantial gain in reliability over other methods. Having established the reliability of our scheme, we therefore move on to describing the distribution of eGEs identified by our scheme, and the predictors of eGE selection by speakers.

GE frequency

For comparability with previous studies, we calculated the rate of GE use per 10,000 words as a means of comparing GE frequency across datasets (although see Pichler 2010, Walker 2010:73–74, and Wagner, Hesson, & Little (2016) for discussion of problems with this measure). The mean rate of GE use in the LCS sample was 25.42 per 10,000 words,[5] and in Toronto it was 28.56, which is a little lower than the rates of around forty per 10,000 words reported for teenagers and young adults in other studies. For example, Stubbe & Holmes (1995) recorded a rate of thirty-nine GEs/10,000 words for eighteen to thirty-four year olds in Wellington, New Zealand, while Cheshire (2007) recorded a rate of 41.12 GEs/10,000 words for fourteen to fifteen year olds in Milton Keynes, UK. For the Fisher corpus, the rate was 42.02, which was more similar to these rates. Despite some disparity in overall frequency, however, the distribution of GE lexical forms was generally similar across the three samples (Table 4).[6]

Of the 1,214 GEs, only a small minority (N = 217, 17.9%) were identified as set-extending eGEs. We also tested the distribution of eGEs and nGEs across corpora to assess whether or not a significantly higher rate of overall GE use was associated with a proportionate increase in eGE use. The eGE use as a proportion of total GE use ranged from 15.7–23.1% across the corpora (see Table 5), suggesting that there may be a proportional difference. A χ^2 test for homogeneity of nGE and eGE distribution by corpus confirms this suspicion ($\chi^2(2) = 7.155$, $p = 0.03$). Multiple two-proportion T-tests between corpora (N = 3 tests; Bonferroni corrected

TABLE 4. *Five most frequent GE lexical forms in LCS, Fisher, and TOR samples by percentage of occurrence within each corpus.*

LCS			Fisher			TOR		
Form	N	%	Form	N	%	Form	N	%
and everything	132	21.8	*and stuff*	59	17.5	*and stuff*	63	23.2
and things like that	98	16.2	*or something*	56	16.6	*or something*	58	21.3
or something	81	13.4	*or whatever*	37	11.0	*or whatever*	43	15.8
or whatever	75	12.4	*and everything*	31	9.2	*and everything*	17	6.3
and stuff like that	52	8.6	*or anything*	22	6.5	*and stuff like that*	12	4.4

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

for multiple comparisons to an alpha value of 0.017) show that this difference is significant for TOR-LCS and TOR-Fisher, but not LCS-Fisher, mirroring the overall frequency results.

The distributional results for total GEs across corpora may reflect, perhaps unsurprisingly, a dialectal difference in the mean GE frequencies between Pennsylvanian US English and Western Ontarian (Torontonian) Canadian English. That being said, the fact that proportional eGE use varied with overall GE use is somewhat unexpected. It suggests that GEs may be serving different functions in Pennsylvanian US English and Torontonian Canadian English. If corroborated with diachronic data, this finding could support a grammaticalization hypothesis for GEs (Cheshire 2007). In other words, it could indicate that GEs are more semantically bleached in US English than in Canadian English. We discuss the grammaticalization hypothesis in more detail in the DISCUSSION section.

On a basic level however, it is difficult to directly interpret these distributional findings. Our coding scheme cannot confirm that the nGE tokens in each corpus did NOT perform set-extending functions. Rather, in each of the 997 nGE cases, the speaker's intended meaning could simply not be determined under the deliberately conservative criteria of our scheme. The 997 nGEs therefore quite likely comprise both (i) GEs that were intended by the speaker and/or understood by the listener to be set-extending, and (ii) GEs that were NOT intended by the speaker and/or understood by the listener to be set-extending. Since this is the case, additional analysis is needed to determine how speakers are using eGEs (vis à vis nGEs), both in terms of GE-internal characteristics and their conditioning linguistic environments. For this purpose, we turn to a multivariate analysis of the linguistic factors associated with eGE production in each corpus.

Multivariate analysis

Once GE tokens had been identified and coded in terms of PRAGMATIC FUNCTION (our dependent variable) as either nGE or eGEs, they were then coded for the independent variables in Table 6. The independent variables include external variables CORPUS and SPEAKER (as a random effect: see the next section), and a number of internal linguistic variables. We derived our coding categories from previously

TABLE 5. *Distribution of eGEs and nGEs.*

	Fisher	LCS	TOR
GE tokens (N)	337	605	272
eGEs	59	95	63
nGEs	278	510	209
% eGE	17.5	15.7	23.1

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

employed schemes, principally Tagliamonte & Denis (2010) and Pichler & Levey (2011). Clarifying examples are given in Table 6.

Variables 4 to 7 were hypothesized to be possible predictors of GE PRAGMATIC FUNCTION (see variable 1), given suggestions in the literature (e.g. Dubois 1992) that variable compositional elements of a GE may be associated with its function (s) in discourse. Like the other variables, 8 and 9 have been employed in previous studies, in which they were intended to serve as indicators of grammaticalization (e.g. Tagliamonte & Denis 2010; Pichler & Levey 2011). As mentioned earlier, Cheshire (2007) suggested that the diversity of GE functions might reflect an ongoing grammaticalization process. Under this theory, although the purportedly

TABLE 6. *Coding categories for quantitative analysis.*

Variable		Levels ('factors')	Clarifying examples
1	PRAGMATIC FUNCTION	eGE, nGE	
2	CORPUS	LCS, Fisher, TOR	
3	SPEAKER	individual speaker names or codes	
4	CONNECTOR	*and*, *or*, zero	*I love bananas **and** stuff like that.* *I love bananas **or** stuff like that.* *I love bananas, ∅ stuff like that.*
5	QUANTIFIER	*no, all, every, some, other, any*, zero	***nothing** like that* ***all** things like that* ***everything** like that* ***something** like that* ***other** things like that* ***anything** like that* *∅ things like that*
6	GENERIC	*stuff, thing, things, where, one*, zero, other	***stuff** like that* ***anything** like that* ***things** like that* ***somewhere** like that* ***someone** like that* *and ∅ all* ***shit** like that*
7	COMPARATIVE	*like, kind*, other, zero	*and stuff **like** that* *and that **kind** of thing* *and things **of that nature*** *and stuff ∅*
8	SYNTAGMATIC LENGTH	long (3 + lexemes), short (1–2 lexemes)	*and stuff like that* *and stuff*
9	SYNTACTIC POSITION	sentence-final/utterance-final, clause-internal/clause-final	*I can get done and everything* *I can get my work and everything done.*

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

prototypical set-marking function is retained for some GEs, others are said to have undergone morphophonological reduction, morphosyntactic decategorization, and semantic bleaching, with punctors representing the final step on this cline (see also Pichler & Levey 2011). In sum, grammaticalized GEs would be short (with respect to SYNTAGMATIC LENGTH), semantically bleached of their extending function, and serving instead a variety of interpersonal and/or textual functions. Such GEs would be more likely to occur in peripheral SYNTACTIC POSITION (sentence- or utterance-final in our taxonomy) since this is a well-known site for discourse phenomena that have undergone semantic-pragmatic bleaching (Brinton 2003:150). We did not test for Cheshire's (2007) other indicator of grammaticalization, namely dissociation between the grammatical category of a GE's referent and the generic noun embedded in its lexical structure. That is—for example, *He was laughing and joking and stuff*—where there is a mismatch between nominal *stuff* and the verbs *laugh* and *joke* to which it refers. Under our coding scheme, however, only eGEs have identifiable referents. Thus we cannot determine whether they are more likely than nGEs to have nominal referents.

We conducted mixed-effects multiple logistic regression analyses using the Rbrul package (Johnson 2009) for the R statistical environment (R Core Team 2013). Some of our original fixed effect factors, such as presence of a COMPARATIVE and LENGTH of the GE naturally interacted: long GEs typically contain comparatives (e.g. *and things of that kind*). We handled this by systematically testing a number of models, each with a different combination of factors. The resulting models were assessed for significant differences using χ^2 tests. Our test models that included COMPARATIVE provided a worse fit to the data than those containing LENGTH. CONNECTOR interacted with both LENGTH and GENERIC due to some empty cells; for example, short GEs without connectors did not occur in the data, and so CONNECTOR was not included in the Rbrul analysis. We also used a random predictor: SPEAKER in analyses of individual corpora. Importantly, in contrast to earlier studies, we use all of these factors to predict the FUNCTION of GEs in our data, rather than to understand the distribution of lexical forms. Recall that this is a key departure from previous work relying on GE form as the sole parameter of GE variability.

LCS and Fisher

For our main analysis (Table 7), we included only the two larger datasets from LCS and Fisher. There was no effect of CORPUS on the distribution of eGEs, that is, the two Pennsylvania corpora are not significantly different with respect to their distribution of eGEs. We tested for the random effect of speaker in a separate run (since each speaker is tied to a unique corpus, they could not be run with CORPUS), and found it to have a significant effect. In other words, there is considerable inter-individual variation in eGE use. The effect of LENGTH was still there, however, even when speaker effects were taken into account. We do not report this in a table for considerations of space.

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

TABLE 7. *Factors influencing the likelihood of eGE use in the Fisher and LCS corpora (i.e. with* PRAGMATIC FUNCTION *as the dependent variable and eGE as the application value).*[a]

Deviance	826.432			
Df	2			
Grand mean 'eGE'	.163			
Factor	Log odds	Tokens (N)	% eGE	Weight
Length				
long (3 + lexemes)	0.321	321	22.4	0.58
short (1–2 lexemes)	−0.321	621	13.2	0.42
Total		942		

[a] Not selected: CORPUS, GENERIC, QUANTIFIER, SYNTACTIC POSITION, STYLE

For the combined Fisher and LCS corpora, LENGTH was the only statistically significant predictor of the extension function, with extending GEs more likely to be long (2 + lexemes) than short. That is, an eGE is more likely to be *and things like that* than *and things.* This may support the claim that length and semantic referentiality are associated, as suggested by the grammaticalization hypothesis.

Indeed, LENGTH is a robust predictor of eGE use within each of the two corpora (Tables 8 and 9). More linguistic predictors of eGE function, however, are operative in LCS than in Fisher. In the LCS corpus (Table 8), the presence of a GENERIC also increased the likelihood of a GE being referential, as did a more 'internal' SYNTACTIC POSITION (i.e. tokens that were not clause-final, utterance-final, or contained in an adjunct). The combination of within-sentence or -utterance position and longer

TABLE 8. *Factors influencing the likelihood of eGE use in the LCS corpus (i.e. with* PRAGMATIC FUNCTION *as the dependent variable and eGE as the application value).*[a]

Deviance	501.249			
Df	6			
Grand mean 'eGE'	.157			
Factor	Log odds	Tokens (N)	% eGE	Weight
Syntactic position				
clause-internal, clause-final	0.398	86	0.256	0.598
sentence/utterance-final	−0.398	519	0.141	0.402
Syntagmatic length				
long (2 + lexemes)	0.344	211	0.209	0.585
short (1 lexeme)	−0.344	394	0.129	0.415
Generic				
stuff	0.275	102	0.255	0.568
no generic	0.215	109	0.193	0.553
thing(s)	−0.490	394	0.122	0.380
Total		605		

[a] Not selected: QUANTIFIER, STYLE; random effect selected: SPEAKER

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

TABLE 9. *Factors influencing the likelihood of eGE use in the Fisher corpus (i.e. with* PRAGMATIC FUNCTION *as the dependent variable and eGE as the application value).*[a]

Deviance	826.432			
Df	2			
Grand mean 'eGE'	.163			
Factor	Log odds	Tokens (N)	% eGE	Weight
Length				
long (2 + lexemes)	0.388	110	25.5	0.596
short (1 lexeme)	−0.388	227	13.7	0.404
Total		337		

[a] Not selected: QUANTIFIER, STYLE, GENERIC, SYNTACTIC POSITION; random effect selected: SPEAKER

SYNTAGMATIC LENGTH may also provide support for the hypothesis that GEs with the 'prototypical' extension function are less grammaticalized than the more utterance-peripheral, semantically bleached GEs. This outcome also supports the qualitative observations of other researchers. Martinez (2011:2466), for example, reported that set-extension appeared to correspond to 'full' (i.e. long) GE forms that occurred 'in the middle of the speaker's turn'. As to why these effects are evident in LCS but not in Fisher, this may be due to the effect of speaker homogeneity. The LCS speakers all attended the same high school, knew one another, and came from the same part of Philadelphia. It should then perhaps be no surprise that they share a number of common constraints on eGEs, which the more geographically diffuse Fisher speakers do not. One might think of the LCS speakers as participants in a Community of Practice (Eckert & McConnell-Ginet 1992; Wenger 1998), where their additional constraints on eGE use emerged within a socially circumscribed subdialect of larger regional and supraregional dialects. In other words, the LCS speakers appear to be subject to the regional constraint of length, while also having their own, specific rules for when to use an unambiguously referential form.

Toronto

By conducting a similar analysis on TOR, a corpus from a different dialect of English with contrastive frequencies of eGE use, we hoped to further test the extent to which constraints on eGE use are uniform across speech communities. Unfortunately, the effect of inter-individual variation (i.e. the random intercept for SPEAKER) was considerable in the nine-speaker TOR sample. In initial runs over the TOR data, only SPEAKER was selected as a significant effect, which suggests that there are no systematic constraints on the pragmatic functions of GEs in TOR—not entirely unexpected given the small size of the dataset. Lowering the threshold of significance from an alpha of .05 to 0.15, however, yields the additional effect of LENGTH (Table 10), hinting at a consistent (albeit statistically insignificant)

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

TABLE 10. *Factors influencing the likelihood of eGE use in the TOR corpus (i.e. with* PRAGMATIC FUNCTION *as the dependent variable and eGE as the application value).*[a]

Deviance	291.836			
Df	3			
Grand mean 'eGE'	.232			
Factor	Log odds	Tokens (N)	% eGE	Weight
Length (at $p < 0.15$)				
long (2 + lexemes)	0.319	56	30.4	0.579
short (1 lexeme)	−0.319	216	21.3	0.421
Total		272		

[a] Not selected: QUANTIFIER, GENERIC, SYNTACTIC POSITION, STYLE; random effect selected: SPEAKER

influence from LENGTH in the same direction as the effect observed in LCS and Fisher.

The question of whether the constraints on eGE use truly are similar across different dialects of English remains to be further tested in other, larger datasets. All the same, we find it striking that eGEs in TOR are likelier to be long than short, just as in LCS and Fisher. This suggests that GEs' referential function is at least partially lexically encoded: when extending sets, speakers do not select from their repertoire of GE forms at random, but are more likely to select GEs that are three lexemes long or longer.

DISCUSSION

The major contribution of this work is to further refine our understanding of how GE use varies across samples through the utilization of novel methodology. In order to address the issue of inter-sample and inter-dialectal variation, we focused on one aspect of Pichler's (2010) call to discourse-pragmatic variation researchers: the need for greater and more careful attention to the function of discourse-pragmatic variables. By creating, testing, and applying an objective method for the identification of set-extending GEs, we brought a new source of data to quantitative, discourse-pragmatic variation studies, that is, a rigorously operationalized concept of the eGE. Using this new measure, we then demonstrated a remarkable similarity in the way speakers functionally distinguish GEs in three geographically contrastive samples. We showed that syntagmatic length was the major predictor of eGE use in Philadelphian US English, Pennsylvanian US English, and Torontonian CA English. Though our findings in the Toronto sample were complicated by small sample size and a large degree of inter-speaker variability, we take the overall results as suggestive evidence that speakers are more alike than different in their deployment of referential GE variants. That being said, our data also illustrated the generalization that smaller, more homogeneous speech communities (such as the Philadelphian high school students in the LCS corpus) are likely to

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

have more constraints on a given variable than more heterogenous speech communities, adding their own flair to broader patterns in their respective regional dialect area. The following subsections: (i) further discuss our observed association between referentiality and length, (ii) connect our findings to an extant hypothesis relating length to semantic-pragmatic interpretation, and (iii) comment on the broader implications of our methodology for future studies on discourse-pragmatic variation.

Associations between GE form and function

An association between GE form (particularly its length) and its referential function has been observed anecdotally in previous studies. Overstreet & Yule (1997:225), for example, note that reduced discourse-pragmatic forms (e.g. *you know what I mean* > *y'know; or anything of that nature* > *or anything*) appear to reinforce intersubjectivity. That is, greater inexplicitness indicates (or implies) greater social closeness and solidarity, signalling that the speakers are socially close enough to leave some things unsaid, including phonetic material. Our quantitative analysis supports these observations: referential eGEs, the GEs that we have presented as less ambiguous and more explicit, are more likely to be long than nGEs. These longer GE forms mostly derive their additional length from comparatives: *of that nature, like that, of that kind*, and so on. They invite the listener to explicitly compare previously mentioned referents with other unmentioned entities in the evoked set. Short GE forms do not explicitly call on the listener to make a mental comparison, and thus may be generally less effective referential devices, making them less preferred for this function. It is of course an open question whether short GEs are dispreferred for referentiality among interlocutors who know each other very well, where, as Overstreet & Yule (1997) point out, a high degree of explicitness may be unnecessary. This is a question for qualitative researchers to investigate on a more micro-level.

Grammaticalization

Our finding that eGEs are more likely than nGEs to be syntagmatically long could be interpreted as tentative support for a hypothesis of GE grammaticalization under which longer GE forms (e.g. *and stuff like that*) become reduced to a shorter form (e.g. *and stuff*) over time, and the shorter forms take on additional discourse functions. Cheshire (2007) also speculated that grammaticalized GEs would occur principally at clause and utterance boundaries, and for the LCS data we found that nGEs were more likely to occur here. Though it should be noted that the factor weight ranges for our results are quite narrow (e.g. range = 16 for LENGTH in the Fisher and LCS corpora; see Table 5), suggesting due caution in interpreting their significance, the patterns demonstrated here are relatively consistent across corpora and relevant factors. Previous quantitative analyses of GE length have yielded mixed results, which may be due to the fact that these analyses did not differentiate between GE

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

function types. Cheshire (2007), Pichler & Levey (2011), and Martinez (2011) all found that shorter forms were generally more frequent than longer forms among young people, which appears to support the idea of ongoing grammaticalization. However, Tagliamonte & Denis (2010) found that both the shorter and longer forms of the GEs in their corpus were stable across speaker age. In the absence of diachronic data (though see Denis 2015) the grammaticalization hypothesis remains a hypothesis. Furthermore, as Dubois (1992) found in her longitudinal Montreal French corpus, and Tagliamonte & Denis (2010:349, Figure 1) found in a meta-analysis of seven studies, GE use is subject to age grading (Sankoff 2005; Wagner 2012), with teenagers and young adults more likely than other age groups to employ GEs. Since the LCS, Fisher, and Toronto datasets comprise only teenagers and young adults, one could interpret our results as reflecting an age grading phenomenon whereby young people use high rates of short nGEs for interpersonal functions, while adults (hypothetically) do not make this length distinction. Finally, the question of whether the (hypothetical) syntagmatic reduction of GEs proceeds across the board or via lexical diffusion could also be explored by looking at long nGEs. Given the present heterogeneity of the nGE category, however, we do not address this question here.[7]

Accountability

Though we assert that our scheme accountably identifies unequivocally referential GEs (i.e. eGEs), we cannot make the same claim for other GE functions. Recall that our scheme does not isolate GE tokens that EXCLUSIVELY extend a set; it only selects GEs that DEFINITELY extend a set. Determining whether any individual eGE also simultaneously performs other discourse functions is beyond the scope of our scheme. Subjective, intuition-based coding may be better suited to untangling this pragmatic nuance. Furthermore, our nGE category is known to be heterogeneous. It includes GEs that perform discourse functions, social functions, and ambiguously referential functions among others. We leave it to future work to objectively circumscribe these remaining functions. This task will almost certainly require a scheme that extends beyond the syntactically derived methods presented here. It may, in fact, be the case that social functions in particular are best treated by qualitative means. Our work suggests that both qualitative and quantitative methods are needed in the study of discourse-pragmatic variation and that their findings can (and should) be mutually informative.

Reliability

Even the most thoughtful examination of the data by one (Cheshire 2007) or two (Tagliamonte & Denis 2010; Pichler & Levey 2011) analysts can be somewhat unreliable and open to subjective interpretation. Acknowledging the caveats with regard to accountability mentioned above, we have nonetheless demonstrated increases in reliability. To our knowledge, there is no sociolinguistic standard for reliability metrics (e.g. Cohen's Kappa values), which are used routinely in the social

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

sciences. We suggest that reliability should be a major concern for variationist researchers studying pragmatic variables/pragmatically constrained variation. Without it, we cannot confidently compare results between studies and thus understand the role of pragmatic variation in either synchronic or diachronic processes.

Moving forward

From an even broader perspective, our article builds on a current trend towards sharing data in sociolinguistics. A number of researchers have, either directly or indirectly, called upon their colleagues to make their datasets available for community use (e.g. Kendall 2008; MacWhinney & Wagner 2010; Kendall & Van Herk 2011). We argue that this push towards collaborative practices should not be limited to data. It should also include protocols that are not currently published in journal articles for reasons of space, such as full coding schemes, metadata, and data processing procedures. Sharing of course occurs informally between colleagues, but is not yet a systematic and more widely expected practice. Sharing these methodologies allows for more productive intra-disciplinary dialogue, giving us a common language by which to compare results and trends across time and geographic space.

CONCLUSION

This study of general extenders in the speech of young women in US and Canadian varieties of English has shown that referential GEs have at least one structural property in common. Across these varieties, unambiguously referential GEs are likelier to be lexically long than their nonreferential counterparts. This result affirms the intuition of Overstreet & Yule (1997), who suggested that longer GEs are more explicit, and thus may more clearly evoke an inferable set. It may also support a theory of GE grammaticalization, although this must be tested using diachronic data.

Our novel method of identifying referential GEs—which we hope will serve as a foundation for further refinement—allowed us to code a large number of tokens for pragmatic function, and to treat pragmatic function as a dependent variable in a logistic regression analysis. To our knowledge, no previous study has used regression to examine the linguistic and external predictors of GE referential function. Having demonstrated this as a possible avenue for researchers in quantitative discourse-pragmatic variation, we hope that comparative synchronic and diachronic analyses of GE referentiality will be undertaken, and that our approach may perhaps be extended to other multifunctional discourse-pragmatic variables. To facilitate replicable future research, we provided the details of our coding method and a demonstration of its inter-rater reliability. We advocate here the development of rigorous methods for identifying pragmatic functions one at a time, gradually building a set of best practices that can be extended to other variables and other functions over time. Like Pichler (2010:584) we hope to 'stimulate serious reflection and

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

discussion amongst discourse variationists in order to generate a GRADUAL progression towards the formulation of a coherent theory of discourse variation and change' (emphasis added).

The question of what speakers might be DOING with the GEs they produce has been the subject of much thoughtful work by both qualitative and quantitative researchers. In the absence of true insights into speaker intentions and listener interpretations, this question may never be fully answered. Nonetheless, our study demonstrates that so long as one function of GEs can be identified in a replicable way, we can know what speakers are doing with some of their GEs at least some of the time. Importantly for sociolinguists, we can then map this information across dialects and across time.

NOTES

[1]We utilize examples from three corpora:

(i) LCS = Language Change and Stabilization corpus (Wagner 2008); I = interview number; speaker names are pseudonyms.
(ii) TOR = Toronto English Archive, followed by corpus code and speaker code.
(iii) Fisher = Fisher English Training Parts 1 and 2, Transcripts (Cieri, Graff, Kimball, Miller, & Walker 2004, 2005), followed by file number and speaker personal identification number (PIN).

[2]In their study of a northern variety of British English, Pichler & Levey (2011) also employed the strategy of isolating punctor GEs from other functional types, but their taxonomy is four-way rather than binary (see Pichler & Levey 2011:450–53 for details; see also the DISCUSSION section).

[3]Pichler & Levey (2011:458) calculate the proportion of punctor tokens for a single GE lexical form, *and that*. They find that *and that* is used more frequently as a punctor than for any of the other functions in their scheme.

[4]We additionally consider category names as potential referents (cf. Dines 1980), though they are apparently rarer than set member type referents. In *fruits like bananas and everything,* for instance, fruits could also count as a referent, meaning that *and everything* would be referential to both bananas and fruit.

[5]One speaker from the LCS was excluded from this calculation. Qualitatively, she appeared to be using GEs at an anomalously high rate compared to other speakers in the same corpus. Since mean frequency is sensitive to outliers, we were concerned that her inclusion would inappropriately inflate the LCS mean. We used Dixon's Q Test (Dixon 1950), a common measure for identifying statistical outliers, to quantitatively test her rate of use relative to the rest of the LCS speakers. The difference between her frequency and that of the next highest frequency GE user was divided by the range of frequency values to calculate a Q score. Based on the Q distribution, the speaker in question was found to be significantly deviant from the rest of the sample and excluded accordingly.

[6]The most frequent GE forms in our corpora are both adjunctive (i.e. contain the connector *and*, as in *and things like that*) and disjunctive (i.e. contain the connector *or*, as in *or whatever*). This may be indicative, as Overstreet (1999) has suggested, of a division of functional labor, with adjunctive forms performing positive politeness, and disjunctive forms performing negative politeness. Cheshire (2007:185–86) further suggests that these politeness functions accrue to GE forms as they grammaticalize. An examination of politeness functions is beyond the scope of this paper and thus we do not consider it further here. For a thorough discussion of the intersection of politeness with GE function see Wagner et al. (2016).

[7]We are grateful to an anonymous reviewer for this suggestion.

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

REFERENCES

Aijmer, Karin (1985). What happens at the end of our utterances? The use of utterance final tags introduced by 'and' and 'or'. In Ole Togeby (ed.), *Papers from the Eighth Scandinavian Conference of Linguistics*, 366–89. Copenhagen: Institut for Philologie.

——— (2002). *English discourse particles: Evidence from a corpus*. (Studies in corpus linguistics 10.) Amsterdam: John Benjamins.

Ball, Catherine, & Mira Ariel (1978). Or something, etc. *Penn Review of Linguistics* 3:35–45.

Brinton, Laurel J. (2003). Historical discourse analysis. In Deborah Schiffrin, Deborah Tannen, & Heidi Hamilton (eds.), *The handbook of discourse analysis*. Oxford: Blackwell.

Buchstaller, Isabelle (2008). The localization of global linguistic variants. *English World-Wide* 29:15–44.

Channell, Joanna (1994). *Vague language*. Oxford: Oxford University Press.

Cheshire, Jenny (1987). Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25:257–82.

——— (2007). Discourse variation, grammaticalization, and stuff like that. *Journal of Sociolinguistics* 11:155–93.

———; Paul Kerswill; Sue Fox; & Eivind Torgersen (2011). Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics* 15:151–96.

Chomsky, Noam (1995). The minimalist program. Cambridge, MA: MIT Press.

Cieri, Christopher; David Graff; Owen Kimball; Dave Miller; & Kevin Walker (2004). *Fisher English training speech part 1 transcripts*. Philadelphia: Linguistic Data Consortium.

———; (2005). *Fisher English training speech part 2 transcripts*. Philadelphia: Linguistic Data Consortium.

D'Arcy, Alexandra (2006). Lexical replacement and the like(s). *American Speech* 81:339–57.

——— (2012). The diachrony of quotation: Evidence from New Zealand English. *Language Variation and Change* 24:343–69.

Denis, Derek (2011). Innovators and innovation: Tracking the innovators of 'and stuff' in York English. In Meredith Tamminga (ed.), *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 39* 17(2). Online: http://repository.upenn.edu/pwpl/vol17/iss2/.

——— (2015). *The development of pragmatic markers in Canadian English*. Toronto: University of Toronto dissertation.

Dines, Elisabeth R. (1980). Variation in discourse and stuff like that. *Language in Society* 9:13–33.

Dixon, W. J. (1950). Analysis of extreme values. *Annals of Mathematical Statistics* 21:488–506.

Dubois, Sylvie (1992). Extension particles, etc. *Language Variation and Change* 4:179–203.

Eckert, Penelope, & Sally McConnell-Ginet (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21:461–90.

Fischer, Kerstin (1998). Discourse particles, turn-taking, and the semantics-pragmatics interface. *Revue de Sémantique et Pragmatique* 8:111–37.

Gwet, Kilem L. (2012). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters*. Gaithersburg, MD: Advanced Analytics.

Hinneburg, Alexander; Heikki Mannila; Samuli Kaislaniemi; Terttu Nevalainen; & Helena Raumolin-Brunberg (2007). How to handle small samples: Bootstrap and Bayesian methods in the analysis of linguistic change. *Literary and Linguistic Computing* 22:137–50.

Jefferson, Gail (1990). List-construction as a task and resource. In George Psathas (ed.), *Interaction competence*, 63–92. Lanham, MD: University Press of America.

Johnson, Daniel Ezra (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3:359–83.

Kendall, Tyler (2008). On the history and future of sociolinguistic data. *Language and Linguistics Compass* 2:332–51.

———, & Gerard Van Herk (2011). Corpus linguistics and sociolinguistic inquiry: Introduction to special issue. Special issue of *Corpus Linguistics and Linguistic theory* 7(1):1–6.

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

Labov, William (2006). *The social stratification of English in New York City*. 2nd edn. Cambridge: Cambridge University Press.
Landis, J. Richard, & Gary G. Koch (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159–74.
Lavandera, Beatriz (1978). Where does the sociolinguistic variable stop? *Language in Society* 7:171–83.
MacWhinney, Brian, & Johannes Wagner (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprächsforschung* 11:154–73.
Martinez, Ignacio M. P. (2011). 'I might, I might go I mean it depends on money things and stuff': A preliminary analysis of general extenders in British teenagers' discourse. *Journal of Pragmatics* 43:2452–70.
Norrby, Catrin, & Joanne Winter (2002). Affiliation in adolescents' use of discourse extenders. In Cynthia Allen (ed.), *Proceedings of the 2001 Conference of the Australian Linguistic Society*. Online: http://www.als.asn.au.
O'Keeffe, Anne (2004). 'Like the wise virgins and all that jazz': Using a corpus to examine vague categorisation and shared knowledge. *Language and Computers* 52:1–20.
Overstreet, Maryann (1999). *Whales, candlelight, and stuff like that: General extenders in English discourse*. Oxford: Oxford University Press.
———, & George Yule (1997). On being inexplicit and stuff in contemporary American English. *Journal of English Linguistics* 25:250–58.
Pichler, Heike (2010). Methods in discourse analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14:581–608.
———, & Stephen Levey (2010). Variability in the co-occurrence of discourse features. In Lynda J. O'Brien & Davide S. Giannoni (eds.), *University of Reading Language Studies Working Papers* 2:17–27.
———, & Stephen Levey (2011). In search of grammaticalization in synchronic dialect data: General extenders in north-east England. *English Language and Linguistics* 15:441–71.
R Core Team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Online: www.R-project.org.
Sankoff, Gillian (2005). Cross-sectional and longitudinal studies in sociolinguistics. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, & Peter Trudgill (eds.), *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society*, vol. 2, 1003–13. Berlin: Mouton de Gruyter.
Schiffrin, Deborah (1994). *Approaches to discourse*. Oxford: Blackwell.
Singler, John V. (2001). Why you can't do a VARBRUL study of quotatives and what such a study can show us. *University of Pennsylvania Working Papers in Linguistics* 7(3):257–78.
Stubbe, Maria, & Janet Holmes (1995). *You know, eh* and other exasperating expressions: An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication* 15:63–88.
Tagliamonte, Sali (2003–2006). *Linguistic changes in Canada entering the 21st century*. Research Grant 410-2003-0005. Ottawa: Social Sciences and Humanities Research Council of Canada.
——— (2005). So who? Like how? Just what? Discourse markers in the conversations of English speaking youth. *Journal of Pragmatics*. 37:1896–1915.
——— (2007–2010). *Directions of change in Canadian English*. Research Grant 410-070-048. Ottawa: Social Sciences and Humanities Research Council of Canada.
——— (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Oxford: Wiley-Blackwell.
———, & Derek Denis (2010). The stuff of change: General extenders in Toronto, Canada. *Journal of English Linguistics* 38:335–68.
Wagner, Suzanne Evans (2008). *Language change and stabilization in the transition from adolescence to adulthod*. Philadelphia, PA: University of Pennsylvania dissertation.
——— (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass* 6:371–82.

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press

———; Ashley Hesson; & Heidi Little (2016). Comparing referential general extender use across registers in American English speech. In Heike Pichler (ed.), *Discourse-pragmatic variation and change in English: New methods and insights*. Cambridge: Cambridge University Press, to appear.
Walker, James (2010). *Variation in linguistic systems*. New York: Routledge.
Ward, Gregory, & Betty J. Birner (1993). The semantics and pragmatics of 'and everything'. *Journal of Pragmatics* 19:205–14.
Wenger, Etienne (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.
Youssef, Valerie (1993). Marking solidarity across the Trinidad speech community: The use of *an ting* in medical counselling to break down power differentials. *Discourse & Society* 4:291–306.

(Received 26 June 2014; revision received 10 January 2015;
accepted 12 March 2015; final revision received 12 April 2015)

https://doi.org/10.1017/S0047404515000603 Published online by Cambridge University Press