Pichler, Heike & Stephen Levey. forthcoming 2011. In search of grammaticalization in synchronic dialect data: General extenders in north-east England. *English Language and Linguistics* 15(3).

In search of grammaticalization in synchronic dialect data: General extenders in north-east England[1]

HEIKE PICHLER & STEPHEN LEVEY
University of Salford & University of Ottawa

(Received 19 March 2010; revised 7 February 2011)

NOTES

[1] The first author gratefully acknowledges the generous support of the Carnegie Trust for the Universities of Scotland which funded the fieldwork undertaken for this project. Earlier versions of this paper were presented at *i-mean* (University of the West of England, April 2009), NWAV 38 (University of Ottawa, October 2009), the Langwidge Sandwidge seminar series (University of Manchester, September 2010) and the Language Variation & Linguistic Theory seminar series (University of Lancaster, December 2010). We would like to thank the audience members for their insightful comments and questions, in particular Ruth Carroll, Jenny Cheshire, Lynn Clark, Derek Denis, David Denison, Willem Hollmann, Paul Kerswill, Miriam Meyerhoff, Anna Siewierska and Elizabeth Traugott. We are also grateful for the detailed comments on previous written versions of the paper made by Karin Aijmer, two anonymous reviewers and editor David Denison. Of course, any remaining errors are our own.

ABSTRACT

In this paper, we draw on a socially stratified corpus of dialect data collected in north-east England to test recent proposals that grammaticalization processes are implicated in the synchronic variability of general extenders (GEs), i.e., phrase- or clause-final constructions such as *and that* and *or something*. Combining theoretical insights from the framework of grammaticalization with the empirical methods of variationist sociolinguistics, we operationalize key diagnostics of grammaticalization (syntagmatic length, decategorialization, semantic-pragmatic change) as independent factor groups in the quantitative analysis of GE variability. While multivariate analyses reveal rapid changes in apparent time to the social conditioning of some GE variants in our data, they do not reveal any evidence of systematic changes in the linguistic conditioning of variants in apparent time that would confirm an interpretation of ongoing grammaticalization. These results lead us to question Cheshire's (2007) recent hypothesis that GEs are grammaticalizing in contemporary varieties of British English. They additionally raise caveats with regard to the assumption that the linguistic conditioning of GE variability in contemporary datasets is the product of change.

1 INTRODUCTION

Hopper & Traugott's (2003: 126) call for reliable statistical studies to diagnose grammaticalization has witnessed a recent increase in studies which apply variationist methodologies to elucidating the processes 'whereby a lexical item or construction in certain uses takes on grammatical characteristics, or through which a grammatical item becomes more grammatical' (Hopper & Traugott 2003: 2). The framework of grammaticalization provides testable hypotheses about the nature and trajectory of changes (decategorialization, desemanticization, etc.) which may be observable in fluid patterns of synchronic dialect variation by virtue of the fact that these changes typically occur as a series of micro-steps (Traugott & Trousdale 2010). The methodological framework of variationist sociolinguistics enables scholars to operationalize these hypotheses as independent linguistic variables conditioning synchronic variation (Poplack 2011; Torres Cacoullos 2011), and to systematically compare the distribution and linguistic conditioning of variant choice across different age groups. This combined approach allows scholars to identify and track grammaticalization pathways in apparent time, as illustrated, for example, in Poplack & Tagliamonte (1999).

Previous attempts to synergize these frameworks in synchronic dialect studies have illuminated the grammaticalization of morpho-syntactic features such as clause-combining constructions and tense-aspect markers (see, for example, Sankoff & Brown 1976; Poplack & Tagliamonte 1999, 2001; Schwenter & Torres Cacoullos 2008). These studies have improved understanding of grammaticalization processes and given rise to important theoretical and methodological refinements: scholars have recognized the need to situate the analysis of grammaticalizing elements in relation to the broader grammatical sub-systems which host them (see, for example, Poplack & Tagliamonte

1996; Tagliamonte & Smith 2006), and have enhanced procedures for identifying and tracking subtle shifts in the linguistic conditioning of variant choice over time (see, for example, Poplack 2011). In addition, scholars have increasingly recognized the value of extending this combined approach to the synchronic study of discourse-level features (see, for example, Tagliamonte & Hudson 1999; D'Arcy 2005).[2]

In this article, we adopt this synthesized approach to investigate recent claims that grammaticalization processes are implicated in the synchronic variability of general extenders (henceforth GEs), i.e., phrase- or clause-final constructions which exhibit extensive variability, as illustrated in (1) and (2) below.[3]

(1) I don't think it's a Berwick word, moggy. You hear it on the telly **and that**, moggy, or cartoons **or whatever**. Moggy.[4]

(2) Everyone in Spittal were fishermen **or whatever**. And all the wives were, you know, doing the nets **and stuff like that**, so that everyone was in Spittal. Just like all the people out at Scremerston were miners **and all the rest of it**.

---

[2] Previously, researchers have utilized the framework of grammaticalization in non-quantitative investigations of discourse-pragmatic change in diachronic corpora (see, for example, Brinton 1996; Traugott 1995, 2003).

[3]The constructions bolded in (1) and (2) have been referred to by various designations (see Cheshire [2007: 156] for an overview). In this paper, we follow Overstreet (1999) and Cheshire (2007) and refer to them as 'general extenders.'

[4] All examples are reproduced *verbatim* from our corpus of interview data collected in Berwick-upon-Tweed, north-east England (see further Section 3.1). Throughout, we use bold typeface for GEs. Where relevant, we use underlining for the constituent over which GEs have scope.

Arguing that frequent GE variants have undergone a set of changes associated with grammaticalization, i.e., phonetic reduction, decategorialization and semantic-pragmatic change, recent studies attribute synchronic patterns of variation in the formal, morpho-syntactic and functional characteristics of GE variants to their differential positioning along a continuum of grammaticalization (Aijmer 2002; Cheshire 2007; Tagliamonte & Denis 2010). Our analysis, which is based on a socially stratified corpus of interview data collected in Berwick-upon-Tweed, north-east England, builds on these recent studies in order to (i) situate patterns of GE variation in this peripheral variety in relation to those uncovered for mainstream varieties of English; (ii) establish whether GEs in this variety are the locus of ongoing change, and, if so, whether grammaticalization is implicated; and (iii) evaluate the extent to which synchronic dialect data can offer a window on the diachronic evolution of GEs.

In Section 2, we review the literature which provides the impetus for the present study, and examine claims that GEs are grammaticalizing in contemporary varieties of English. We then detail our method for tracking grammaticalization in our corpus in Section 3. In Section 4, we present the results of our analysis and situate them in relation to previous research. We then interpret these results and discuss their implications in Section 5, before concluding and pointing to areas of further study in Section 6.

2 PREVIOUS RESEARCH ON GES

2.1 *Social and geographical variation*

GE variability has been investigated in many varieties of English, including the English of England (Aijmer 2002; Cheshire 2007; Denis ms.; Levey ms.); Scotland (Macaulay 1991); Ireland (O'Keeffe 2004); America (Overstreet 1999; Overstreet & Yule 1997, 2002); Canada (Tagliamonte & Denis 2010); New Zealand (Britain 1992; Stubbe & Holmes 1995); and Australia (Dines 1980; Norrby & Winter 2002). These studies have correlated variability in GE usage with a number of social factors, including age, sex, socio-economic class, and locality.

The most consistent sociolinguistic patterns have been uncovered for age. Although overall mean frequencies vary, young people typically make the greatest use of GEs across varieties (Denis ms.; Stubbe & Holmes 1995; Tagliamonte & Denis 2010; but see Stenström et al. 2002: 89). This may be a result of either age-grading (Dubois 1992) or ongoing change (Stubbe & Holmes 1995: 83).

Significant patterns of sex differentiation have been reported for New Zealand English (Britain 1992), with females using GEs more often than males. Other studies, however, have either noted no marked differences in rates of use associated with sex (Cheshire 2007; Norrby & Winter 2002), or observed that the effect of sex on variant choice is weaker than that of other social variables (Tagliamonte & Denis 2010). An interaction of sex with socio-economic class is noted by Stubbe & Holmes (1995: 77) for New Zealand English: young middle-class females and young working-class males are the most prolific users of GEs.

Britain (1992), Dines (1980) and Macaulay (1991) found that GEs are more commonly used by working-class than middle-class speakers in New Zealand, Australian and Scottish English. Cheshire (2007) reported the absence of consistent social class differences in the frequency of GE usage amongst adolescents in England.

However, she found that the use of specific variants is sensitive to social class membership: while middle-class adolescents favour *and stuff* and *and things*, working-class adolescents favour *and (all) that*. The strong association of *and that* with young working-class speech is also reported in Levey's (ms.) study of London pre-adolescents.

Cross-variety comparisons reveal interesting correlations of GE usage with locality. Overall, *and that* is strongly associated with New Zealand and British English varieties (Britain 1992; Cheshire 2007; Denis ms.; Levey ms.), while *and stuff* is strongly associated with North American English varieties (Overstreet 2005; Overstreet & Yule 1997; Tagliamonte & Denis 2010). However, the high frequency of *and stuff* amongst middle-class adolescents in Cheshire's (2007) data from Reading, Milton Keynes and Hull, three towns in south-eastern and northern England, as well as its noted frequency in Denis's (ms.) data from York, a city in northern England, suggest that this variant is on the increase in urban varieties of British English.

2.2 *Formal and functional variation*

Dines (1980: 18) describes the structural configuration of GEs using the following schema: AND/OR [PRO FORM] (LIKE THAT), where PRO FORM includes generic referents such as *anything, stuff, something, things*, and parentheses indicate optional lexical material. The connectors *and/or* provide the basis for a structural sub-division between two main types of GEs: adjunctive variants (e.g. *and stuff, and things*) and disjunctive variants (e.g. *or something like that*). Use of a connector, while common, is not obligatory (e.g. *things like that, anything like that*).

Early studies of GEs were generally more concerned with characterizing their function rather than their form. Dines (1980: 22), for example, emphasized their set-

marking function, claiming that GEs are prototypically used to 'cue the listener to interpret the preceding element as an illustrative example of a more general case'. This usage is exemplified in (3), where *or whatever* evokes the fully lexicalized category of 'business people' (an interpretation supported by the explicit mention of the category in the immediately following clause), and (4), where *and stuff* evokes an *ad hoc* or 'fuzzy categor[y]' (Overstreet 1999: 43) related to concepts such as 'speaking politely', 'speaking standard English', etc.

(3) Berwick was run by what they call freemen and burgesses. And these people were the, generally the wealthy people in the town. Or the tradesmen, the goldsmiths. Or the silversmiths, or the butchers **or whatever**. The people who've had businesses.

(4) If it's like the bank, I mean, of course you have to speak clearly **and stuff**.

The highly context-driven interpretation of GEs has prompted scholars to increasingly recognize that GEs can be exploited for a range of interactional purposes which extend beyond set-marking, including interpersonal functions such as signalling politeness or solidarity (Aijmer 2002; Cheshire 2007; Norrby & Winter 2002; Overstreet 1999), and textual functions such as signalling turn-exchange or transitions (Aijmer 2002; Cheshire 2007). Aijmer (2002: 217–218) and Cheshire (2007: 183) suggest that the multifunctionality of GEs in synchronic data may be the result of gradual semantic-pragmatic changes over time, with interpersonal and textual functions developing from the putatively original set-marking function. These developments are commonly situated within a larger set of changes associated with grammaticalization.

2.3 *GE variability and grammaticalization*

Much recent research on GEs has focused on assessing the extent to which GEs have undergone the processes that are collectively associated with grammaticalization (Heine 2003: 579): phonetic reduction, characterized by the loss of phonetic substance (Aijmer 2002: 227; Erman 1995: 145; Overstreet 1999; Overstreet & Yule 1997); decategorialization, involving the loss of morpho-syntactic characteristics of source forms, and their extension beyond their originally defining morpho-syntactic contexts (Aijmer 2002: 217–218; Erman 1995; Overstreet 1999: 27; Overstreet & Yule 1997); as well as semantic-pragmatic change, whereby forms become semantically bleached and acquire new pragmatic meanings over time (Aijmer 2002: 217–218; Erman 1995; Overstreet & Yule 1997). To track the grammaticalization of GEs in synchronic datasets of British and Canadian English, Cheshire (2007) and Tagliamonte & Denis (2010) operationalized the aforementioned processes as independent variables for quantitative analysis.

Drawing on hypotheses about the evolution of short GE variants from longer variants (e.g. *or something* < *or something like that*), Cheshire (2007) equated phonetic reduction with loss of lexical material (e.g. *like that*), and argued that in the British English varieties of Reading, Milton Keynes and Hull, some variants (e.g. *and everything*, *and that*) were more advanced in terms of phonetic reduction than others (e.g. *or something*). While Denis (ms.) claimed to have uncovered evidence of increasing phonetic reduction for variants containing the generics *stuff*, *things* and *something* in his data from York (UK), Tagliamonte & Denis (2010) did so only for *something* variants in their Toronto data. In both York and Toronto, GEs containing

other generic nouns generally exhibited considerable stability in apparent time between long and short variants.

Decategorialization was gauged by Cheshire (2007), Levey (ms.) and Tagliamonte & Denis (2010) in terms of mismatches between the morpho-syntactic and semantic properties of the generic in GEs (e.g. *thing(s), stuff*) and those of the NPs to which GEs are appended, as well as the nature of the antecedent phrase (e.g. VP, AdjP). This diagnostic is illustrated in (5)–(7):

(5) I have seen it frae time to time on maybe a news item **or something like that**.

(6) It's usually all on history programmes **and stuff**.

(7) They're no being aggressive **or anything**.

In (5), the properties of the generic, *something*, in the GE ([+COUNT] [+SINGULAR] [+INANIMATE]) match those of the preceding head noun *news item*. By contrast, in (6) and (7), there is a mismatch between the properties of the generic noun in the GE and the antecedent phrase: the generic noun *stuff* in (6) is not appended to an expected mass noun, but to a plural inanimate count noun, *history programmes*; in (7), the GE is appended to an AdjP, exemplifying the putative extension of GEs to new contexts (Himmelmann 2004). Contexts exemplified in (5) have been construed in the literature as the original host context for GEs, while those in (6) and (7) are seen to instantiate later stages in the grammaticalization process.

Applying these measures, Cheshire (2007), Levey (ms.) and Tagliamonte & Denis (2010) found that individual GE variants exhibited varying degrees of decategorialization. However, Tagliamonte & Denis's (2010) apparent time study

revealed little evidence to support claims that GEs were becoming increasingly decategorialized. With the exception of *and things*, the distribution of GEs across different morpho-syntactic contexts was found to be relatively stable in their Toronto data.

To track the semantic-pragmatic development of GEs in their corpora, Cheshire (2007) and Tagliamonte & Denis (2010) first coded each token for loss of referential meaning. Reasoning that the source meaning of GEs is to implicate a more general set (Dines 1980), they operationalized the absence of a set-marking function as an indication of semantic bleaching. A token implicating a lexicalized superordinate category, (8), was thus taken to be less advanced in terms of semantic bleaching than a token not implicating a superordinate category, (9).

(8) You never hear Berwick making the news because of attacks on people **and stuff like that**. [larger category of 'crime' or 'violence']

(9) We found that eh stuff **and that** was dearer for a while.

In Cheshire's (2007) data, short GEs were reported to have undergone some degree of semantic bleaching, with *and that* affected the most. Tagliamonte & Denis (2010) found minimal evidence of semantic bleaching of GEs in their Toronto data. Although *stuff* variants were more likely to be void of a set-marking function amongst younger speakers, patterns of variability in apparent time did not provide any unequivocal evidence of ongoing change.

Finally, in order to determine the degree of pragmatic change in GE usage, Cheshire (2007) and Tagliamonte & Denis (2010) coded all GE tokens for their co-

occurrence with other discourse-level features in the ambient discourse. Arguing that GEs in the earlier stages of pragmatic change will tend to be reinforced by the presence of additional discourse-level features performing similar functions, they interpreted low levels of co-occurrence to be an indication of advanced pragmatic change. While co-occurrence patterns were generally low in Cheshire's (2007: 185) British English data, they tended to be higher in Tagliamonte & Denis's (2010: 357) Toronto English data. Crucially, Tagliamonte & Denis (2010) did not uncover any evidence to suggest that co-occurrence patterns are decreasing in apparent time.

In concluding this section, we note that Cheshire's (2007) proposal that GEs in contemporary varieties of English are grammaticalizing must be treated cautiously since it is based on the analysis of adolescent speech only, and is not confirmed by Tagliamonte & Denis's (2010) apparent time study of GE variability in Toronto English.

3 DATA AND METHODOLOGY

In this section, we introduce the methodology we implement to probe the existence and characterize the potential direction of grammaticalizing changes in our corpus. In doing so, we critically review some of the measures that have previously been invoked by researchers to detect and track ongoing grammaticalization of GEs in synchronic data. Among the methodological refinements that we introduce are revised criteria for operationalizing syntagmatic length of GEs as a measure of grammaticalization as well as a novel approach to assessing semantic-pragmatic change.

3.1 *Data*

Our study is based on a 260,000-word corpus collected between 2003 and 2005 in Berwick-upon-Tweed, a small market town of just over 15,000 inhabitants located in the far north-east of England.[5] The data were gathered using the interview protocol designed for the *Survey of Regional English* to solicit information about informants' use of local dialect words and attitudes towards their locality and dialect (Llamas 2007). Various fieldwork techniques were utilized to mitigate potential effects of the Observer's Paradox (Labov 1972: 209), including conducting interviews with self-selected pairs in informants' homes (Labov 1972: 210). The quality of the data suggests that such effects were successfully attenuated, as reflected by the rich repository of vernacular features characteristic of casual and unmonitored speech styles (Labov 1972: 85–99), such as non-standard *was*, relativizer *what*, negative concord and double modal constructions.

As shown in Table 1, the sample includes 36 speakers who are equally stratified across sex and three emically defined age cohorts reflecting shared life stages: teenagers and young adults in temporary non-career employment or full-time education, and co-habiting with their parents (17–23); adults in full-time salaried employment with dependent children and their own household (27–48); retired people with no dependent children (60–81). They represent a socially homogeneous group of working-class speakers, as determined via a combination of traditional social class indicators (housing, education, occupation) and speakers' own assessment of their social class membership.

[insert Table 1 about here]

[5] Our word count includes false starts, truncations, filled pauses and minimal response particles, but excludes the interviewer's contributions.

The design of our study affords new insights into GE variability: the focus on the geographically peripheral variety of Berwick English (BwE) complements the recent focus on mainstream varieties (Cheshire 2007; Denis ms.; Levey ms.; Tagliamonte & Denis 2010), and the inclusion of three age groups transcends the limitations of studies which have inferred change in GEs from investigation of one age group only (Cheshire 2007; Levey ms.).

3.2 *Methodology*

3.2.1 *Definition of the variable and circumscription of the variable context*

Dines's (1980: 23) functionally-based definition of the variable context, which assumes that GEs share the unitary function of marking the preceding element as a member of a set, is not supported by recent studies that foreground the functional versatility of GEs (see Section 2.2 above). Given the difficulties in defining membership of the set of GEs by function, we adopt Pichler's (2010) structure-based conceptualization of discourse variables. We define GEs as semi-fixed constructions which prototypically share the following schematic pattern: (connector) (modifier) (generic noun/pro-form) (similative) (deictic), where parentheses indicate the optionality of individual components, as illustrated in Table 2. We also include in the variable context less prototypical GE variants which appear to be more fixed in terms of their structural configuration such as *and all the rest of it, and so on (like that), or whatever it is, type of thing*. The inclusion of these constructions in the variable context is justified on the grounds that they occur in the same syntactic environments as more prototypical variants, and perform a similar range of functions to these. This approach to defining

the variable context allows us to study the whole spectrum of GE variants under the umbrella of one variable, while still catering for the variable's inherent multifunctionality and potential semantic-pragmatic changes over time.

[insert Table 2 about here]

Having defined the variable context this way, we exhaustively extracted all GE constructions from the data. We excluded tokens whose surface form could not be disambiguated, tokens where insufficient context hampered unambiguous data interpretation, and tokens that occurred in quoted speech, repair sequences, or contexts that did not unequivocally support GE interpretations. In total, 783 GE tokens were retained for analysis.

3.2.2 *Coding*

To investigate GE variability and grammaticalization in BwE, each GE token in the dataset was coded for its variant form (see Appendices 1–3 for a full breakdown of individual variants). Additionally, each GE token was coded for speaker sex, one of the three age cohorts as well as for syntagmatic length, decategorialization and semantic-pragmatic change.

In operationalizing syntagmatic length as a diagnostic of grammaticalization, we depart from previous approaches which appeal to diachronic processes of phonetic reduction (Cheshire 2007). We do this for two reasons. Firstly, the derivation of short variants from longer variants (e.g. *and things* < *and things like that*) is questionable from a historical perspective (see, for example, Carroll [2008] who reports that short

variants such as *and such* feature among the earliest attestations of GEs). Secondly, as acknowledged by Tagliamonte & Denis (2010), scholars of grammaticalization habitually characterize phonetic reduction in terms of segmental loss concomitant with shifts in morpheme boundaries (e.g. *going to* > *gonna*), rather than in terms of the loss of whole morphemes. In this study, we hypothesize that if morphologically lighter forms tend to be more grammaticalized than morphologically heavier forms (Hopper & Traugott 2003: 106; see also Givón's [1990] 'quantity principle'), then grammaticalization may be evidenced in apparent time by an increase in the number of short GE variants relative to long ones (see also Levey ms.). To operationalize syntagmatic length as an independent variable, we therefore coded all GE tokens for whether they were short (1 or 2 lexemes) or long (3+ lexemes).

To assess decategorialization, we follow the lead of Cheshire (2007) and Tagliamonte & Denis (2010) and examine the collocation of variants with expected NPs, where GE generics match the morpho-syntactic features of preceding NPs (10); cases where there is a mismatch between preceding NPs and appended GEs in terms of these features (11); and cases where GEs are appended to non-nominal phrases, i.e., VPs, AdjPs, numerical expressions, clauses and quoted speech (12). Tokens that occur after elicited dialect words, as in (13), were coded separately because their categorization is not straightforward.[6]

[6] When informants follow an elicited dialect word with a GE, it is not clear whether the GE refers to a NP, implicating a more general set of dialect words ('and other words like this'), or to different word classes depending on the semantics of the elicited word (e.g. AdjP 'and other adjectives like this' when following 'knackered' in [13]).

(10) (a) It must depend on like the weather **and stuff**.

(b) I think they have got to go through an act of parliament **or something**.

(11) (a) You never hear Berwick making the news because of attacks on people **and stuff like that**.

(b) See the dolphins here **and everything** sometimes, you know.

(12) (a) You never hear of people getting murdered **and things like that** in Berwick, you know.

(b) I think it's probably homely **and everything**

(c) It's the same thing that's in it **and everything**.

(d) And the media goes, 'Gerard K.O.' **and that**.

(13) I probably say eh knackered **or something like that**.

We used a slightly modified protocol to gauge the decategorialization of GE variants such as *and that* and *or whatever* which do not contain a generic noun that can be co-referenced with an antecedent NP. We coded these variants for whether they occurred with NPs, (14), or in extended contexts with non-nominal phrases, (15).[7]

(14) (a) I don't know whether it's the adult sheep **or whatever**.

(b) And she's a good artist **and that**.

(15) (a) If you wanted to be anonymous **or whatever**.

(b) And eh we split up **and that**.

[7] A small number of GEs were not coded for decategorialization because their reference was ambiguous.

Our investigation of semantic-pragmatic change diverges from recent approaches which operationalized this parameter as if it consisted of two independent processes, and inferred pragmatic shift from the co-occurrence of GEs with other discourse-level features (Cheshire 2007; Tagliamonte & Denis 2010). These approaches are based on the questionable assumptions that co-occurrence patterns are functionally motivated, and that they constitute a straightforward metric of semantic-pragmatic change (see Cameron & Flores-Ferrán [2004] and Pichler & Levey [2010] for evidence to the contrary). With others, we believe that the interplay of referential and pragmatic/procedural meanings is intrinsic to semantic-pragmatic change during the incipient stages of change (Company 2006; Heine 2002), and that set-marking GEs are inherently intersubjective because they appeal to common frames of knowledge by implicitly evoking a more general set (O'Keeffe 2004). Any operationalization of semantic-pragmatic change should reflect this. Thus, we submit that in the course of semantic-pragmatic change, the set-marking meaning of GEs gradually recedes while their intersubjective and other pragmatic/procedural meanings increasingly come to the fore (see Hopper & Traugott 2003: 94).

To model this process, we coded every GE token in the data for whether or not it performed (a) the referential function of implicating a more general set (see examples (8)–(9) in Section 2.3), and (b) one or more interpersonal and/or textual functions, as illustrated in (16)–(18). In (16), Janet's use of *or something* signals her uncertainty about whether the preceding expression 'ulterior motives' is the one she intended, which is also reflected in the false-starts and truncated question at the beginning of her turn. In (17), Patrick uses *and things like that* as a means to secure his hold on the floor and prevent his interview partner Carol, who repeatedly signals her desire to contribute

to the discussion ('yeah,' 'yeah, like' and in-breath), from taking the floor. Finally, in (18), Alicia's *and stuff like that* serves as a retrospective hedge on the preceding description of her own accent as well as a turn-yielding signal, an effect which is achieved by the production of the GE in a separate tone unit and with a falling intonation contour.[8]

(16) Janet: Alterna- what was the what's the word I'm trying to? Em, ulterior motives **or something**. Somebody that's being particularly nice to you but they're no really. They're just being leery.

(17) Patrick: It's the way people pronounce their [rs]. Their ts. **And**

Carol: [Yeah,] Yeah, like

Patrick: = **things like that**. [And (.)] the way they skip words. They shorten (.)

Carol: [*(h)*]

Patrick: = they skip a word and they make it really short.

Carol: Like the the (.) like (.) say the had a letter on the end, they would like miss that out.

(18) Alicia: But I quite like my accent. It's got touches of different sort of maybe a

---

[8]In examples (16)–(18) and in Table 3 below, we have used the following transcription conventions:

[ ]	overlap	(.)	short pause
=	turn continuation	.	final intonation contour
-	truncation	,	continuing intonation contour
(h), (.h)	inbreath, outbreath	?	rising intonation contour

Informant names are pseudonyms; HP is the interviewer.

bit of American when I speak, and a bit of Scottish. [**And stuff like**

HP: [Mhm.] Is there

Alicia: = **that**.]

HP: = bits of Geordie in it as well?

Tokens following dialect words and measurements were not coded for their semantic and pragmatic meanings: with elicited dialect words the general set had been invoked by the interviewer (e.g. 'Are there any dialect words for *tired*?'); with measurements it is unclear whether they are indeed intended to evoke a category of possible measurements (see Cheshire 2007: 176). Tokens which implicate a more general set by invoking shared frames of reference were coded as not performing a pragmatic function since, as noted above, their appeal to shared knowledge is intrinsic to these tokens' set-marking function.

Drawing on current hypotheses about the unidirectionality and regularity of semantic-pragmatic change (truth-conditional/content/non-subjective > non-truth-conditional/procedural/(inter)subjective) (Traugott & Dasher 2002: 40), we used the coding protocol detailed above to elaborate the taxonomy outlined in Table 3.The taxonomy models the continuous shifts in the semantic-pragmatic evolution of GEs along a cline of grammaticalization, similar to that outlined in Heine (2002) for grammaticalizing material more generally. At Stage 0, tokens perform the core referential function of GEs, i.e., they mark the preceding referent as a member of a more general set. Stage 1 gives rise to the interpersonal and textual meanings inherent in the source form which co-exist with the original referential meaning, with the layering of meanings reflecting the gradualness of semantic-pragmatic change. At Stage 2, the

referential meaning fades away and pragmatic meanings are strengthened. Finally, at Stage 3, in line with Traugott & Heine's (1991) implicational hierarchy of the type semantic > pragmatic > less semantic-pragmatic, earlier semantic-pragmatic meanings fade away, and GEs come to function as 'punctors' (Vincent & Sankoff 1992), i.e. tokens which are entirely desemanticized and which serve to punctuate or bracket units of discourse (see also Macaulay 1985; Cheshire 2007).

[insert Table 3 about here]

4 RESULTS

4.1 Distributional results

Figure 1 reveals the uneven distribution of GE variants in our data. The most frequent variant, *and that*, constitutes almost one third of all GE tokens in the data. *Or something, or whatever* and *and things like that* are the next most frequent variants, constituting 9%, 7% and 5% of the data respectively. Together with *and that*, these variants account for just over half of all GE tokens in the dataset. *And stuff, or something like that, and everything* as well as *and things* are also comparatively frequent, although none accounts for more than 3-4% of the data. The remaining 33% of GE tokens comprises 86 variants, each instantiated fewer than 20 times in the corpus. Inspection of age patterns (not shown here) confirms that *and that* and *or something* are the most frequent variants in each age cohort. Among the other frequent variants, we notice the following trends: a slight decline of *and everything* and stark decline of *or whatever* among the youngest speakers; a decline of *things* variants among middle and young speakers; and a gradual rise of *and stuff* from older to younger speakers. With

regard to the infrequent variants in the data, apparent time distributions (not shown here) reveal that older male speakers have a much larger inventory of these than the other speaker groups (see further Table 8 below).

The highly fragmented nature of the variable context is by no means an isolated finding. Denis's (ms.) corpus of York English contains 87 GE variants which are also unevenly partitioned between a limited number of relatively frequent variants and a much larger number of relatively infrequent ones. Cross-variety comparisons reveal additional similarities: *and that* and *or something* are consistently among the most productive variants in the British English varieties examined to date (Aijmer 2002; Cheshire 2007; Denis ms.; Levey ms.). Considering the working-class composition of our speaker sample, the elevated frequency in our data of *and that* also bolsters previous claims that this variant is indexical of British English working-class speech (Cheshire 2007; Levey ms.).

[insert Figure 1 about here]

Figure 2 compares the normalized distribution of GEs across sex and age. Males in every age cohort, particularly the youngest, use GEs more frequently than females, and younger speakers as a whole use GEs more than older ones. In the absence of a significant increase in GE usage between the old and middle groups, higher rates of GE use by the young group are suggestive of an age-graded effect, which has been widely reported for other varieties, both in Britain and beyond (Denis ms.; Tagliamonte & Denis 2010).

[insert Figure 2 about here]

The formal distribution of GE variants in Figure 1 above reveals that the more frequent variants in our data tend to be syntagmatically short. In order to examine whether this tendency is indicative of possible ongoing change associated with grammaticalization, we conducted paired-sample t-tests. These tests reveal whether differences in the mean frequency of short and long GE variants are statistically significant within each age group. Comparison of the results across age groups will then reveal whether there are changes in apparent time in the distribution of long versus short variants.[9] Table 4 reveals that short *something* variants increase in apparent time, concomitant with a reduction in the use of long *something* variants. Only in the youngest age cohort, however, does the difference in syntagmatic length achieve statistical significance (see Denis [ms.] for similar results). By contrast, variants containing generic *things* reveal no propensity to be short, as evidenced by the fact that long *things* variants are more common in each age cohort (see Cheshire [2007] and Denis [ms.] for similar results). No significant differences emerge for *stuff* variants, although short constructions are more common than longer ones, particularly in the youngest group. In the case of *everything* variants, short constructions are more frequent than long ones in the old and middle cohorts (young speakers are omitted due to sparse

[9] The comparisons are confined to frequent variants containing a generic. *And that* as well as GEs containing *whatever* are excluded from this part of the analysis: with *and that*, it is unclear which (if any) of the longer constructions could serve as a direct point of comparison; with *whatever* variants, there is an insufficient number of long variants (e.g. *or whatever it is*) for viable statistical analysis.

data), but the difference of means is not significant. Summarizing, with the possible exception of *something* variants, our analysis yields little evidence to suggest that long variants are gradually being supplanted by short variants (see Tagliamonte & Denis [2010] for similar results).

[insert Table 4 about here]

We take these findings as a provisional indication that stability, rather than change, may be a key determinant of patterns of GE variability in our data. In order to substantiate this hypothesis, we next examine the underlying structure of surface patterns of GE variability. This structure emerges from detailed multivariate analyses of the social and linguistic factors conditioning variant choice when all factors are considered simultaneously.

4.2 *Results of multivariate analyses*

The highly skewed formal distribution of GEs in our data presents a particular challenge for statistical analysis. Only a small minority of GE variants occur frequently enough to permit viable multivariate analyses of individual variants. An analysis which focused exclusively on the proportionally most dominant variants would necessarily entail the neglect of a large portion of the variable context containing the bulk of the remaining variants. To circumvent this problem, we configure the data in such a way that allows us to conduct multivariate analyses of all variants, including frequent and infrequent ones. At least two reasons justify this approach. Firstly, it allows us to contextualize any potential changes instantiated by the most frequent GE variants in relation to the larger

linguistic sub-system in which they are embedded (see Weinreich et al. [1968] on the 'embedding problem'), *a fortiori* because change does not occur in a systemic vacuum (Traugott & Trousdale 2010: 29). Secondly, it allows us to explore whether low-frequency GE variants may be just as susceptible to grammaticalization processes as frequent variants, as recently hypothesized by Hoffmann (2005) and Brems (2007) with regard to other variables.

With these considerations in mind, we configure the ensuing multivariate analyses as follows. Table 5 focuses on the proportionally most dominant variant, *and that*. This variant is sufficiently frequent in the data to allow independent multivariate analyses to be conducted for each age group. This enables us to track any potential trajectories of change in apparent time. In Table 6, we focus on the next most frequent variants in the data: *or something* and *or whatever.* As neither variant is quantitatively sufficient to conduct independent multivariate analyses for each age group, we examine these variants across all groups combined. While this configuration of the data does not afford an apparent time window on the linguistic conditioning of variant selection, it does enable us to ascertain whether each variant is increasing or decreasing in apparent time when age is factored into the analysis as an independent variable. Table 7 presents the results of an apparent time analysis of other frequent GE variants containing a generic referent. Because of the relatively low numbers of individual variants, these variants are analyzed in the aggregate. Infrequent variants (N<20) are also treated as an aggregate, and the results of the apparent time analysis are displayed in Table 8. While aggregated sets of variants underpinning the analyses in Tables 7 and 8 may obscure the behaviour of individual variants, comparison among the analyses presented in Tables 5–8 validates

each as an important means of building up a composite picture of patterns of variation and change in the BwE GE cohort.

To model the variability in the data, and to identify the relative contribution to variant choice of the independent variables (also known as 'factor groups') described in Section 3.2 (age, sex, syntagmatic length, decategorialization, semantic-pragmatic change), we make use of the stepwise multiple regression procedure in GoldVarb X (Sankoff et al. 2005). We derive three lines of evidence from the multivariate analyses (Poplack & Tagliamonte 2001: 93–94): (i) the statistical significance of individual factors (at the 0.05 level); (ii) the *range*, which indicates the relative strength of factor groups included in the analysis; and (iii) the *constraint hierarchy*, which is the ordering of factor weights within a factor group offering a window on the grammar underlying the surface variability.[10] In each of the multivariate analyses shown in Tables 5–8, the input value indicates the overall probability of variant occurrence. Factor weights indicate the likelihood of occurrence of variants in the contexts listed on the left-hand-side of the tables. Factor weights >.50 favour the selection of the variant (or aggregates of variants) in the environment in question; factor weights <.50 disfavour variant selection. KO indicates a KnockOut value, i.e., instances where GoldVarb cannot calculate factor weights because the variant is realized either 0 or 100 per cent of the time in a given context within a factor group. Factor weights enclosed in square brackets are non-significant. The *range* enables us to rank the magnitude of individual

[10] Because smaller sample sizes are less likely to achieve thresholds of statistical significance, we construe the constraint hierarchy as a window on the underlying structure of variability even when a factor group is not selected as statistically significant (see also Poplack & Tagliamonte 2001:93).

effects with respect to one another: the greater the range, the greater the effect of the factor group on variant choice. The em-rule indicates that a factor was not included in the analysis due to insufficient data. We make use of these numerical formalisms to (a) determine the relative contribution of social and linguistic factors to variant choice; and (b) ascertain whether systematic comparisons of variable structure across the age cohorts reveal any evidence of ongoing change implicating grammaticalization.

Table 5 presents the results for *and that*. The input values for the three age groups indicate that this variant is more frequent among older and particularly younger speakers than among speakers from the middle age group, a pattern we will return to in Section 5 below. The effect of speaker sex appears to have been substantially reweighted in apparent time: we witness a mild but non-significant tendency for women to use *and that* more than men in the oldest group, but this effect is neutralized in the middle group. Only in the youngest cohort is sex a significant determinant of variant choice, with young males clearly favouring *and that*. The range values show that decategorialization is highly significant in all three age groups. The constraint hierarchies indicate stable patterns of variability in apparent time: *and that* is consistently favoured with nominals. The factor weights for semantic-pragmatic change in Table 5 (as well as those in Tables 6–8) suggest that the meaning of GEs cannot be reduced to a set-marking function (Stage 0), routinely assumed in the literature to be the 'core' or primary function of GEs. On the contrary, multifunctionality is an intrinsic feature of the use of GEs in our data, as evidenced by their probability of occurrence with other stages of our functional taxonomy (see also Tables 6–8). More specifically, the patterns in Table 5 show that *and that* is consistently favoured by all three age groups as a marker encoding interpersonal/textual functions (Stage 2), and is

overwhelmingly favoured as a desemanticized punctor (Stage 3). In fact, *and that* is the only variant that occurs at Stage 3 in the old and young age groups, as indicated by the KO values. In sum, the constraint hierarchies for all linguistic factor groups contrast with those for social factor groups in remaining essentially stable in apparent time, vitiating any inference of ongoing grammaticalization for the most frequent GE variant in BwE.

[insert Table 5 about here]

Table 6 displays the results for *or something* and *or whatever.* The factor weights for age reveal that *or something* is increasing in frequency with decreasing age. By contrast, *or whatever* peaks in the middle group but recedes dramatically in the youngest group where it is almost non-existent. The two variants also differ with regard to semantic-pragmatic meaning: this factor group is not selected as significant for *or something* but contributes a strong effect on *or whatever*. *Or whatever* exhibits a greater propensity than *or something* to occur in contexts associated with set-marking and interpersonal/textual functions (Stage 1). There are, however, also similarities in the patterning of the two variants: both are favoured by females and with nominal antecedents. While we cannot establish from these results whether the linguistic conditioning of *or something* and *or whatever* is changing in apparent time, the patterns revealed here demonstrate that the frequency of *or something* is increasing in apparent time, and that there is some degree of social and functional differentiation between the two variants.

[insert Table 6 about here]

Comparison across age of the input values for the aggregated set of other frequent variants in Table 7 reveals that this group of variants occurs most often in the oldest cohort. Other frequent variants are consistently more likely to occur in female than male speech, and the effect of sex increases in apparent time. With regard to linguistic factors, the significant preference for long variants instantiated in the oldest group is reversed in the middle group (where the difference is non-significant), and is neutralized in the youngest group. Decategorialization makes a strong contribution in all three age groups (see the elevated *range* values): other frequent variants are consistently favoured with NPs, suggesting that they are not undergoing categorial shift. Finally, semantic-pragmatic change is not selected as significant for any age group. The constraint hierarchies are largely parallel across age, with the exception of the youngest speakers, where the most favouring factor is Stage 2. As in the case of *and that,* the picture that emerges from the results for the aggregate of other frequent variants is one which depicts changes in apparent time in the social conditioning of variability but shows no changes in linguistic conditioning.

[insert Table 7 about here]

Turning finally to the infrequent variants, Table 8 reveals a number of important similarities as well as contrasts with those shown in Table 7. Judging from the input values, infrequent forms, when considered in the aggregate, are less likely to be used by the youngest cohort than the middle and older cohorts. Infrequent variants are

consistently associated with males, although not significantly so with young speakers. In terms of linguistic conditioning, syntagmatic length makes the most significant contribution to the occurrence of infrequent variants: these are predominantly long in all three age groups. The results further reveal quantitative evidence of decategorialization. In the oldest group, infrequent variants are highly favoured in non-nominal contexts. However, the fact that this effect is non-significant in the middle group, and is all but neutralized in the young group, is inconsistent with what would be expected if gradual decategorialization were underway. Nor can the results for semantic-pragmatic change be construed as evidence of change in progress: constraint hierarchies are parallel across all age groups, with variants most likely to occur at Stage 1. Once again, these patterns provide no evidence for ongoing linguistic change, but show clear evidence of ongoing social change in the GE cohort: older (male) speakers in this community use a broader inventory of GE variants than younger speakers.

[insert Table 8 about here]

Comparison of the results in Tables 5–8 offers some insight into the relationship between frequency and patterns of variability in our data. The linguistic conditioning of variability differs quite markedly across frequent and infrequent variants, and furnishes little support for construing frequent variants as an analogical model for less frequent variants (see Hoffmann's recent claims [2005: 195] on the interplay between frequency and analogy in grammaticalization).

5 DISCUSSION

The apparent time analyses presented in Section 4 revealed substantive evidence of structured variability in GE usage in BwE. However, apparent time comparisons of the constraint hierarchies for linguistic factors turned up no compelling evidence of grammaticalization in progress. Even in cases where differences in the significance and direction of linguistic effects obtain between age cohorts, the resultant patterns remain inconsistent with any inference of change in progress. With none of the multivariate analyses furnishing evidence of *ongoing* grammaticalization, it remains to be determined whether the synchronically stable patterns of GE variability in our data – and, for that matter, in other synchronic datasets (e.g. Tagliamonte & Denis 2010) – are in fact the product of grammaticalization processes that may have been operative at an earlier stage of the language, predating the time-span in our corpus.

A number of arguments can be marshalled in favour of the hypothesis that stable synchronic patterns of GE variation preserve the effects of diachronic grammaticalization processes. As shown in Section 4, all the linguistic measures specifically operationalized in our study to track grammaticalization revealed considerable variability in terms of the formal, morpho-syntactic and functional properties of GEs. The patterning of individual GE variants along these linguistic measures is consonant with previous research which has shown that the various processes constituting grammaticalization may not unfold concurrently (Croft 1990: 244), and that none of these processes, once initiated, necessarily has to go to completion (Traugott & Trousdale 2010: 28). The net result of arrested sets of changes that have not evolved in synchrony may be a 'ragged and incomplete subsystem that is not evidently moving in some identifiable direction' (Hopper & Traugott 2003: 131). If we hypothesize that the stable linguistic conditioning of GE variants in our synchronic

data is the reflex of such incomplete and asynchronous grammaticalization processes, then Hopper and Traugott's observation may explain how individual GE variants in our data came to be synchronically fossilized at different points along different grammaticalization clines. Consider for example *and that*, which is favoured in extended semantic-pragmatic contexts that are semantically bleached, yet disfavoured in extended morpho-syntactic contexts following non-nominal referents.

Furthermore, inter-variant differences in the linguistic patterning of GEs across factor groups are consistent with previous research which has shown that different variants of a grammaticalizing construction do not 'march [...] through the [grammaticalization] changes in lockstep' (Bybee & Torres Cacoullos 2009: 214). If we assume that the stable linguistic conditioning of GE variants in our synchronic data is the reflex of non-uniform grammaticalization processes, then Bybee & Torres Cacoullos's observation may explain why different GE variants, or different aggregates of variants, pattern differently across independent linguistic variables. Consider for example *or something* and *or whatever*: in the community as a whole, the former is weakly preferred in Stage 0 of our functional taxonomy, while the latter is strongly disfavoured in this environment.

While stable patterns of GE variability may in theory be compatible with a grammaticalization scenario, such a scenario awaits confirmation from diachronic analyses. Without an appropriate real-time benchmark, reconstruction of the diachronic transitions that have given rise to contemporary patterns of GE variability, as well as inferences that these patterns are the result of change, must remain speculative. We foreground this point because a number of recent studies (e.g. Cheshire 2007; Tagliamonte & Denis 2010) have paid insufficient attention to key methodological

requirements for ratifying the existence and directionality of change, and uncritically assumed that synchronic patterns of GE variation can provide a window on their evolution.

A crucial step in the empirical establishment of change involves discriminating authentic changes from cases of inherent variability. As Poplack & Levey (2010) point out, because variation is a precondition for change, but does not necessarily equate with change in and of itself (see also Weinreich et al. 1968: 188), variation may easily be mistaken for change. A good illustration of this point can be found in patterns of variation which persist for considerable lengths of time without being actively engaged in any known trajectory of change: synchronically productive variables such as (t/d)-deletion and (ing) are two well-known cases of stable variability that are deeply rooted in the history of the English language (Romaine 1985). Without diachronic evidence, we cannot rule out the possibility that variation in GEs represents a similar case of longstanding variability. Another issue to be considered is that even if we assume that contemporary patterns of GE variation are ultimately the result of a sequence of changes associated with grammaticalization (or even changes independent of grammaticalization), any changes may have unfolded centuries ago. If this is the case, then the temporal span included in conventional apparent time analyses will be inadequate to identify and chart the full progression of such changes, which are typically slow and protracted (Heine 2002: 95; Lichtenberk 1991: 195).

The caveats discussed above, while not sufficient to refute the grammaticalization hypothesis, certainly suggest a critical reassessment. For example, Tagliamonte & Denis's (2010: 362) recent suggestion that cross-dialectal differences in patterns of GE variability 'may reflect different degrees of grammaticalization or even different

grammaticalization paths' must be interpreted cautiously when the historical record is not available to elucidate the evolution of GEs. Joseph (2004: 60) counsels against reconstructing grammaticalization pathways using synchronic data when the 'range of details of the development and unfolding of a form through time' is incomplete or unknown. This problem needs to be addressed in future research: suitable diachronic data need to be located to track the historical evolution of GEs.

In stark contrast to the stability uncovered in the linguistic conditioning of GE variants, our analysis revealed strong evidence of instability in the social conditioning of variant choice. The results indicate that GEs are prime candidates for the marking of social differentiation. Because GEs in our data reveal no compelling evidence of ongoing grammaticalization, increased rates of use by the young (male) speakers in the sample cannot be interpreted unequivocally as a change in progress but may instead be indicative of age-grading (Dubois 1992). Of particular interest with regard to age-differentiated patterns are the input values for *and that* (see Table 5 above), which show striking resemblance to Downes's (1998: 224) U-curve distribution: in contrast to speakers from the middle age group, speakers from the old and young age groups share similarly elevated frequency rates for this variant. This pattern may be suggestive of speakers retreating from the use of *and that* as they enter the *marché linguistique* (Bourdieu & Boltanski 1975) before picking it up again as they grow older. Whether this behaviour is motivated by the increasing association in this community of *and that* with a male working-class sociolect awaits further investigation.

Our analysis also revealed dramatic changes in the frequency of individual GE variants in contemporary BwE (e.g. *or whatever* rapidly decreases between the middle and young age groups; *and stuff* sharply increases in the youngest group; *things* variants

decline across age). These patterns mirror those reported elsewhere (Denis ms.; Tagliamonte & Denis 2010), and support Labov's (1982: 76) observation that scholars need to look beyond the linguistic conditioning to the social embedding of language in order to fully understand the renewal of individual variants. While social and internal changes in discourse frequently tend to operate in unison (e.g. Tagliamonte & D'Arcy 2007; Tagliamonte 2008), the results presented here suggest that – with some discourse variables at least – social changes in variant choice may occur despite stable linguistic conditioning.

6 CONCLUSION

Our study of GEs in a peripheral variety of north-east England has extended the geographical focus of variationist research on GEs in British English. Our findings broadly resemble those obtained by Cheshire (2007), Denis (ms.), Levey (ms.) and Tagliamonte & Denis (2010) for mainstream British and Canadian English varieties with regard to the highly differentiated nature of the variable context; the propensity for individual variants to wax and wane in apparent time; and the recruitment of specific variants to mark social differences. In addition, the inclusion in our analysis of an apparent time component has offered an important check on Cheshire's (2007) foundational claims that GEs are grammaticalizing in contemporary British English. Our rigorous quantitative analysis did not uncover any systematic incremental patterns in apparent time that would enable us to confirm Cheshire's (2007) hypotheses and characterize variability in terms of ongoing grammaticalization. As we have indicated above, these findings have important analytical and methodological ramifications for

the interpretation of recent work on GE variability as well as the design of future studies of GEs and discourse variables more generally.

A much-needed extension to current variationist work on GEs is the incorporation of a real-time component to increase the time-depth of data analysed. The exploration of diachronic surrogates of oral data may enable us to ascertain the extent to which GEs have been implicated in change, and to determine the role of grammaticalization in any trajectories of change uncovered. The social embedding of GE variability, which is subject to rapid changes within a compressed time frame (see also Tagliamonte & Denis 2010), is an additional key area for further investigation both within and across varieties. Close inspection of the social constraints on variant choice in BwE indicates that *and that* may be in the process of becoming the default variant among young (male) speakers. Future research will show whether the variant *and stuff*, which is increasing to varying degrees in our peripheral variety and Cheshire's (2007) and Denis's (ms.) mainstream varieties, will become a serious contender among the youngest generation.

Crucial to elucidating patterns of GE variation and change in synchronic and diachronic datasets is the development of an empirically accountable analytical framework which caters for GEs' extensive formal variability as well as their inherent multifunctionality. In this paper, we built on Cheshire's (2007) foundational framework for studying GEs in several pivotal ways. Firstly, our operationalization of syntagmatic length in terms of the variation between morphologically 'heavier' and morphologically 'lighter' variants allowed us to side-step the problems associated with conceptualizing this measure of grammaticalization in terms of diachronically unsubstantiated phonetic reduction processes. Secondly, our elaboration of a replicable taxonomy for assessing the functional properties of GEs enabled us to systematically investigate semantic-

pragmatic change in apparent time, and corroborated the importance of function in accounting for GE variability and discourse-pragmatic variability more generally. Finally, we departed from previous investigations of GEs which have tended to privilege the most productive GE variants at the expense of less frequent ones which often constitute the bulk of the data. Our inclusion of the entire set of GE variants in the variable context enabled us to probe previously uninvestigated differences in the patterning of high- and low-frequency variants. Further cross-varietal research, judiciously informed by a diachronic perspective, will allow scholars to resolve some of the outstanding issues we have raised in connection with the role (if any) of grammaticalization in patterns of variation and change in GEs.

Authors' addresses:

Heike Pichler

School of Humanities, Languages & Social Sciences

University of Salford

Maxwell Building

Salford M5 4WT

UK

h.pichler@salford.ac.uk

Stephen Levey

Department of Linguistics

University of Ottawa

70 Laurier Avenue East

Ottawa K1N 6N5

Canada

slevey@uottawa.ca

APPENDIX 1: Full inventory of adjunctive GE variants

VARIANT	N	%
and that	246	31.4
and things like that	38	4.9
and stuff	34	4.4
and everything	28	3.6
and things	25	3.2
and all that	19	2.4
and stuff like that	13	1.7
and everything else	10	1.3
and all the rest of it	8	1
and so on	7	0.9
and elsewhere	4	0.6
and one thing another	4	0.6
and whatever	4	0.6
and all	3	0.4
and all this	3	0.4
and places like that	3	0.4
and different things	2	0.3
and something	2	0.3
and so on like that	2	0.3
and all that kind of stuff	1	0.1
and all that sort of thing	1	0.1
and all them	1	0.1
and all them places	1	0.1
and all these big shops	1	0.1
and anything	1	0.1
and anything like that	1	0.1
and different things like that	1	0.1
and else	1	0.1
and everything else on	1	0.1
and everywhere	1	0.1
and just things like that	1	0.1
and people like that	1	0.1
and physical stuff like that	1	0.1
and that sort of thing	1	0.1
and them	1	0.1
and them different parts	1	0.1
and things like this	1	0.1

and this and that	1	0.1
and what have you	1	0.1
and what not	1	0.1
	476	60.8

APPENDIX 2: Full inventory of disjunctive GE variants

VARIANT	**N**	**%**
or something	68	8.8
or whatever	54	7.0
or something like that	29	3.8
or anything	18	2.3
or anything like that	7	0.9
or so	5	0.7
or somewhere	5	0.7
or somewhere like that	4	0.6
or wherever	4	0.6
or things like that	3	0.4
or anywhere	2	0.3
or anybody else	1	0.1
or any kind of fishing thing	1	0.1
or anyone else	1	0.1
or anywhere like that	1	0.1
or however	1	0.1
or nothing	1	0.1
or nothing like that	1	0.1
or nowt like that	1	0.1
or owt like that	1	0.1
or someone like that	1	0.1
or something or other	1	0.1
or something to that effect	1	0.1
or stuff like that	1	0.1
or that	1	0.1
or whatever it is	1	0.1
or whatever it’s called	1	0.1
or whatever it was	1	0.1
or whatever like that	1	0.1
or whatever the number is	1	0.1
or whatever you call it	1	0.1
or whatever you call them	1	0.1
or whatever you come frae	1	0.1

or whenever	1	0.1
or whoever	1	0.1
	223	28.5

APPENDIX 3: Full inventory of GE variants without connectors

VARIANT	**N**	**%**
things like that	14	1.8
something like that	13	1.7
sort of thing	13	1.7
whatever	12	1.5
anything like that	6	0.8
kind of thing	6	0.8
that sort of thing	4	0.6
etcetera	3	0.4
all that sort of thing	2	0.3
all them places	1	0.1
anything else	1	0.1
anywhere like that	1	0.1
nothing like that	1	0.1
places like that	1	0.1
somewhere like that	1	0.1
sort of type	1	0.1
teams like that	1	0.1
those sort of places	1	0.1
type of thing	1	0.1
type thing	1	0.1
	84	10.7

REFERENCES

Aijmer, Karin. 2002. *English discourse particles: Evidence from a corpus* (Studies in Corpus Linguistics 10). Amsterdam: John Benjamins.

Bourdieu, Pierre & Luc Boltanski. 1975. Le fétichisme de la langue. *Actes de la recherche en sciences sociales* 4, 2–32.

Brems, Lieselotte. 2007. The grammaticalization of small size nouns. Reconsidering frequency and analogy. *Journal of English Linguistics* 35(4), 293–324.

Brinton, Laurel J. 1996. *Pragmatic markers in English. Grammaticalization and discourse function* (Topics in English Linguistics 19). Berlin: Mouton de Gruyter.

Britain, David. 1992. Things and that in Porirua: An analysis of set marking tags. Notes from departmental seminar. Wellington Linguistics Seminar Series.

Bybee, Joan & Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization: How the particular and the general interact in language change. In Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley (eds.), *Formulaic language*, 2 vols, vol. 1, *Distribution and historical change* (Typological Studies in Language 82), 187-218. Amsterdam: John Benjamins.

Cameron, Richard & Nydia Flores-Ferrán. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1), 41–65.

Carroll, Ruth. 2008. Historical English phraseology and the extender tag. *Selim* 15, 7–37.

Cheshire, Jenny. 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11(2), 155–93.

Company, Concepcion Company. 2006. Zero in syntax, ten in pragmatics: Subjectification as syntactic cancellation. In Angeliki Athanasiadou, Costas

Canakis & Bert Cornillie (eds.), *Subjectification: Various paths to subjectivity* (Cognitive Linguistics Research 31), 375–97. Berlin: Mouton de Gruyter.

Croft, William. 1990. *Typology and universals* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

D'Arcy, Alexandra. 2005. Like: Syntax and development. Unpublished PhD dissertation, University of Toronto.

Denis, Derek. ms. Grammaticalization of general extenders in York English. Unpublished manuscript, University of Toronto.

Dines, Elizabeth R. 1980. Variation in discourse – 'and stuff like that'. *Language in Society* 9(1), 13–31.

Downes, William. 1998. *Language and society* (Cambridge Approaches to Linguistics). 2nd ed. Cambridge: Cambridge University Press.

Dubois, Sylvie. 1992. Extension particles, etc. *Language Variation and Change* 4(2), 179–203.

Erman, Britt. 1995. Grammaticalization in progress: The case of *or something*. In Inger Moen, Hanne Gram Simonsen & Helga Lødrup (eds.) *Papers from the fifteenth Scandinavian conference of Linguistics, Oslo, 13-15 January 1995,* 136–47. Oslo: Department of Linguistics, University of Oslo.

Givón, Talmy. 1990. *Syntax.* vol. 2: *A functional typological introduction.* Amsterdam: John Benjamins.

Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization* (Typological Studies in Language 49), 83–101. Amsterdam: John Benjamins.

Heine, Bernd. 2003. Grammaticalization. In Joseph & Janda (eds.), 575–601.

Himmelmann, Nikolaus P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In Walter Bisang, Walter, Nikolaus Himmelmann & Björn Wiemer (eds.) *What makes grammaticalization? A look from its fringes and components* (Trends in Linguistics 158), 21–42. Berlin: Mouton de Gruyter.

Hoffmann, Sebastian. 2005. *Grammaticalization and English complex prepositions. A corpus-based study* (Routledge Advances in Corpus Linguistics). London: Routledge.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Joseph, Brian D. 2004. Rescuing traditional (historical) linguistics from grammaticalization theory. In Olga Fischer, Muriel Norde & Harry Perridon (eds.) *Up and down the cline-The nature of grammaticalization* (Typological Studies in Language 59), 45–71. Amsterdam: John Benjamins.

Joseph, Brian D. & Richard D. Janda (eds.). 2003. *The handbook of historical linguistics*. Oxford: Blackwell.

Labov, William. 1972. *Sociolinguistic patterns.* Oxford: Blackwell.

Labov, William. 1982. Building on empirical foundations. In Winfred P. Lehmann & Yakov Malkiel (eds.) *Perspectives on historical linguistics*, 17–92. Amsterdam: Benjamins.

Levey, Stephen. ms. General extenders and grammaticalization: Insights from London preadolescents. Submitted to *Applied Linguistics.*

Lichtenberk, Frantisek. 1991. On the gradualness of grammaticalization. In Traugott & Heine (eds.), 41–80.

Llamas, Carmen. 2007. A new methodology: Data elicitation for regional and social language variation studies. *York Papers in Linguistics* 8, 138–63.

Macaulay, Ronald. 1985. The narrative skills of a Scottish coal miner. In Manfred Görlach (ed.), *Focus on Scotland* (Varieties of English around the World G5), 101–24. Amsterdam: John Benjamins.

Macaulay, Ronald. 1991. *Locating dialect in discourse: The language of honest men and bonnie lasses in Ayr*. Oxford: Oxford University Press.

Norrby, Catrin & Joanne Winter. 2002. Affiliation in adolescents' use of discourse extenders. *Proceedings of the 2001 conference of the Australian Linguistic Society.* http://www.arts.uaw.edu.au/LingWWW/als 01/proceedings.html (18 December 2010).

O'Keeffe, Anne. 2004. 'Like the wise virgins and all that jazz': Using a corpus to examine vague categorization and shared knowledge. *Language and Computers* 52(1), 1–26.

Overstreet, Maryann. 1999. *Whales, candlelight, and stuff like that: General extenders in English discourse.* Oxford: Oxford University Press.

Overstreet, Maryann. 2005. And stuff *und so*: Investigating pragmatic expressions in English and German. *Journal of Pragmatics* 37(11), 1845–64.

Overstreet, Maryann & George Yule. 1997. On being inexplicit and stuff in contemporary American English. *Journal of English Linguistics* 25(3), 250–8.

Pichler, Heike. 2010. Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14(5), 581–608.

Pichler, Heike & Stephen Levey. 2010. Variability in the co-occurrence of discourse features. *Language Studies Working Papers* 2, 17–27.

Poplack, Shana. 2011. A variationist perspective on grammaticalization. In Heiko Narrog & Bernd Heine (eds.), *Handbook of grammaticalization*, 209-224. Oxford: Oxford University Press.

Poplack, Shana & Stephen Levey. 2010. Contact-induced grammatical change. In Peter Auer & Jürgen Erich Schmidt (eds.) *Language and space. An international handbook of linguistic variation.*vol. 1: *Theories and methods*, 391-418. Berlin: Mouton de Gruyter.

Poplack, Shana & Sali A. Tagliamonte. 1996. Nothing in context: Variation, grammaticalization and past time marking in Nigerian Pidgin English. In Peter Baker (ed.) *Changing meanings, changing functions: Papers relating to grammaticalization in contact languages*, 71–94. Westminster: University of Westminster Press.

Poplack, Shana & Sali A. Tagliamonte. 1999. The grammaticization of 'going to' in (African American) English. *Language Variation and Change* 11(3), 315–342.

Poplack, Shana & Sali A. Tagliamonte. 2001. *African American English in the diaspora: Tense and aspect* (Language in Society 30). Oxford: Blackwell.

Romaine, Suzanne. 1985. The sociolinguistic history of t/d deletion. *Folia Linguistica Historica* 2, 25–59.

Sankoff, David, Sali A. Tagliamonte & Eric Smith. 2005. Goldvarb X. A multivariate analysis application. Department of Linguistics, University of Toronto, Canada. http://indivdiual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm (18 December 2010).

Sankoff, Gillian & Penelope Brown.1976. The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language* 52, 631–66.

Schwenter, Scott & Rena Torres Cacoullos. 2008. Defaults and indeterminacy in temporal grammaticalization: The 'perfect' road to perfectives. *Language Variation and Change* 20(1), 1–39.

Stenström, Anna-Brita, Gisle Andersen & Ingrid Kristine Hasund. 2002. *Trends in teenage talk* (Studies in Corpus Linguistics 8). Amsterdam: John Benjamins.

Stubbe, Maria & Janet Holmes. 1995. *You know, eh* and other exasperating expressions: an analysis of social and stylistic variation in the use of pragmatic particles in a sample of New Zealand English. *Language and Communication* 15(1), 63–88.

Tagliamonte, Sali A. 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics* 12(2), 361–94.

Tagliamonte, Sali A. & Alexandra D'Arcy. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19(2), 199–217.

Tagliamonte, Sali A. & Derek Denis. 2010. The *stuff* of change: General extenders in Toronto, Canada. *Journal of English Linguistics* 38(4), 335–68.

Tagliamonte, Sali A. & Rachel Hudson. 1999. Be like et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3(2), 147–72.

Tagliamonte, Sali A. & Jennifer Smith. 2006. Layering, competition and a twist of fate: Deontic modality in dialects of English. *Diachronica* 23(2), 341–80.

Torres Cacoullos, Rena. 2011. Variation and grammaticalization. In Maria Diaz-Campos (ed.), *The handbook of Hispanic sociolinguistics*, 148-167. Oxford: Wiley-Blackwell.

Traugott, Elizabeth Closs. 1995. The role of the development of discourse markers in a theory of grammaticalization. Paper presented at ICHL XII, Manchester.

Traugott, Elizabeth Closs. 2003. Constructions in grammaticalization. In Joseph & Janda (eds.), 624–47.

Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change* (Cambridge Studies in Linguistics 17). Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs & Bernd Heine. 1991. Introduction. In Traugott& Heine (eds.), 1–14.

Traugott, Elizabeth Closs & Bernd Heine (eds.). 1991. *Approaches to grammaticalization*, vol. 2*: Focus on types of grammatical markers* (Typological Studies in Language 19). Amsterdam: John Benjamins.

Traugott, Elizabeth Closs & Graeme Trousdale. 2010. Gradience, gradualness and grammaticalization. How do they intersect? In Elizabeth Closs Traugott & Graeme Trousdale (eds.), *Gradience, gradualness and grammaticalization* (Typological Studies in Language 90), 19-44. Amsterdam: John Benjamins.

Vincent, Diane & David Sankoff. 1992. Punctors: A pragmatic variable. *Language Variation and Change* 4(2), 206–16.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfried Lehmann & Yakov Malkiel (eds.) *Directions for historical linguistics*, 95–188. Austin: University of Texas Press.

young (17-23)		middle (27-48)		old (60-81)	
male	*female*	*male*	*female*	*male*	*female*
6	6	6	6	6	6

Table 1. Speaker sample

(connector)	(modifier)	(generic noun/ pro-form)	(similative)	(deictic)
and	different	things	like	that
and	different	things		
and		things	like	that
		things	like	that
and		things		
and	physical	stuff	like	that
and		stuff	like	that
		stuff	like	that
and	all that kind of	stuff		
and		stuff		
and	all			that
and	all			
and				that
or		something	like	that
		something	like	that
or		something		
or		whatever	like	that
or		whatever		
		whatever		

Table 2. Prototypical structural pattern of GE variants

stage	function	illustration	interpretation
Stage 0	set-marking (contingent on intersubjectivity)	Daniel: But it was nice to (.) y- nice to think somebody was there [making] your meals, and you could go in HP: [mhm] yeah Daniel: = <u>if something was wrong</u> and you <u>hurt your</u> HP: yeah Daniel: = <u>knee</u> or <u>blacked your eye</u> **or something like that**.	The GE implicates a larger category of 'injuries' or 'accidents'.
Stage 1	set-marking **and** interpersonal/textual	HP: What do you consider the local football derby to be? Godfrey: Well Berwick, it'd be Berwick versus somebody. It's probably <u>Brechen</u>. I should imagine it would be HP: mhm Godfrey: = <u>Brechen</u> because they seem to have a bit of a rivalry wi them for some reason. <u>Or Queen of the South</u>. **Somewhere like that**. I honestly couldn't tell HP: mhm Godfrey: = you. Divn't follow it.	The GE implicates a larger category of 'potential local football rivals.' It also functions as a hedge to signal Godfrey's uncertainty vis-à-vis the accuracy of the provided information as well as a means to signal his desire to close the topic. This interpretation is supported by Godfrey's following acknowledgment that he lacks knowledge of local football, which constitutes an additional disclaimer as well as a legitimate warrant for not sustaining the topic.
Stage 2	interpersonal/textual	HP: But don't you think that the Scots get a better deal? Theodore: Well, th- their rates **and that**, everything's cheaper. [The water rates] and everything. When my niece HP: [mhmmhm] Theodore: = told me what she paid for her water, you know. What we pay here. I think it's terrible like, HP: mhm Theodore: = you know.	The GE does not implicate a larger set of 'things that are cheaper in Scotland than in Northumberland', as indicated in the subsequent discourse which focuses squarely on the topic of 'rates'. The GE functions in the textual domain to foreground a new discourse-entity and to introduce a new topic which is pursued in the remainder of Theodore's turn.
Stage 3	punctor devoid of	HP: Although it's actually in your opinion more Geordie	The GE does not implicate a larger

	referential & pragmatic meanings	you would [still] prefer the Scots. Why? Gabriel: [yeah] *(h)* Just cos e:h some of my family's Scottish **and that**, and I don't really like getting called a Geordie.	category of different nationalities for Gabriel mentioned previously in the interview that his ancestors are without exception from Scotland. The GE serves to punctuate the discourse.

Table 3. Taxonomy of semantic-pragmatic change of GEs

GE variant	Old	Middle	Young
short *something*	**1.5**	**2.4**	**4.9**
long *something*	**1.9**	**1.2**	**1.1**
Difference of means	*t=-0.5, df=11, p=0.63*	*t=0.4,df=11,p=0.69*	***t=3.5,df=11,p=0.01*
short *things*	**1.1**	**0.6**	**1.2**
long *things*	**3.4**	**3.2**	**3.8**
Difference of means	*t=-1.8,df=11, p=0.09*	***t=-2.4,df=11,p=0.03*	*t=-1.5,df=11,p=0.15*
short *stuff*	**0.5**	**1.2**	**2.8**
long *stuff*	**0.3**	**0.8**	**0.9**
Difference of means	*t=0.9,df=11,p=0.39*	*t=0.4,df=11,p=0.67*	*t=1.5,df=11,p=0.16*
short *everything*	**1.3**	**1.2**	N/A
long *everything*	**0.8**	**0.2**	N/A
Difference of means	*t=0.6,df=11,p=0.56*	*t=2.0,df=11,p=0.07*	N/A

Table 4. Results of paired-sample t-tests for distribution of variants by generic and syntagmatic length (bolded figures indicate frequency of variants per 10,000 words; N/A indicates insufficient data; ** indicates statistical significance)

	OLD			MIDDLE			YOUNG	
Input	.49			.25			.56	
Total N	102			43			101	
	FW	%		FW	%		FW	%
Sex								
female	[.56]	38		[.48]	18.1		**.29**	20.7
male	[.48]	30.9		[.51]	19.7		**.63**	52.9
range							*34*	
Decategorialization								
nominal	**.70**	72		**.69**	52.1		**.72**	81.9
non-nominal	**.22**	23.6		**.37**	22.5		**.31**	44.6
range	*48*			*32*			*41*	
Sem.-pra. change								
Stage 0	**.54**	35.6		**.50**	17.6		**.48**	35.4
Stage 1	**.35**	20		**.34**	10		**.40**	28
Stage 2	**.65**	47.2		**.66**	28.9		**.65**	52.5
Stage 3	KO	100		**.97**	85.7		KO	100
range	*30*			*63*			*25*	

Table 5. Three independent multivariate analyses of the contribution of factors to the probability of *and that* by age

	OR SOMETHING			**OR WHATEVER**	
Input	.09			.08	
Total N	68			54	
	FW	%		FW	%
Age					
old	**.40**	5.4		**.39**	5.4
middle	**.52**	8.9		**.65**	14.7
young	**.61**	12.6		—	—
range	*21*			*26*	
Sex					
female	**.59**	12.3		**.58**	9.1
male	**.46**	7		**.46**	5.9
range	*13*			*12*	
Decategorialization					
expected NP	**.66**	17.8		**.58**	11.5
other NP	**.62**	15.5		—	—
non-nominal	**.37**	6.2		**.42**	7.5
range	*29*			*16*	
Sem.-pra. change					
Stage 0	[.54]	9.5		**.23**	1.4
Stage 1	[.50]	8		**.75**	12.8
Stage 2	[.45]	6.8		**.49**	5.1
Stage 3	—	—		—	—
range				*52*	

Table 6. Multivariate analyses of the contribution of factors to the probability of *or something* and *or whatever*

	OLD			MIDDLE			YOUNG	
input	0.34			0.28			0.28	
Total N	73			40			41	
	FW	*%*		*FW*	*%*		*FW*	*%*
Sex								
female	[.58]	32.9		**.64**	34.0		**.82**	43.7
male	[.47]	24.0		**.44**	19.0		**.32**	7.1
range				*22*			*50*	
Length								
short	**.43**	20.0		[.55]	27.3		[.50]	19.5
long	**.61**	36.1		[.44]	19.3		[.50]	19.0
range	*18*							
Decategorialization								
expected NP	**.49**	35.0		**.58**	35.7		**.67**	45.0
other NP	**.77**	65.0		**.75**	54.8		**.75**	54.4
non-nominal	**.33**	22.7		**.31**	14.8		**.38**	20.0
range	*44*			*44*			*37*	
Sem.-pra. change								
Stage 0	[.58]	30.0		[.63]	34.1		[.53]	25.0
Stage 1	[.48]	22.5		[.48]	21.8		[.42]	17.5
Stage 2	[.39]	16.7		[.42]	17.9		[.56]	26.8
Stage 3	—			[.35]	14.3		—	
range								

Table 7. Three independent multivariate analyses of the contribution of factors to the probability of other frequent variants by age (other frequent variants = *and things like that, and stuff, or something like that, and everything, and things*)

	OLD			MIDDLE			YOUNG	
input	0.33			0.36			0.22	
Total N	103			88			70	
	FW	*%*		*FW*	*%*		*FW*	*%*
Sex								
female	**.36**	21.5		**.38**	27.8		[.43]	22.8
Male	**.55**	36.9		**.56**	44.7		[.54]	31.6
range	*19*			*18*				
Length								
short	**.32**	16.7		**.24**	14.9		**.29**	10.3
Long	**.81**	63.9		**.88**	80.7		**.94**	81.0
range	*49*			*64*			*65*	
Decategorialization								
NP	**.45**	29.5		[.48]	36.8		[.49]	24.6
non-nominal	**.64**	47.2		[.54]	42.3		[.52]	27.7
range	*19*							
Sem.-pra. change								
Stage 0	**.48**	31.7		[.47]	39.2		**.46**	24.6
Stage 1	**.62**	45.3		[.52]	43.8		**.65**	41.3
Stage 2	**.36**	22.2		[.50]	42.2		**.35**	16.9
Stage 3	—	—		—	—		—	—
range	*26*						*30*	

Table 8. Three independent multivariate analyses of the contribution of factors to the probability of infrequent variants (N<20) by age

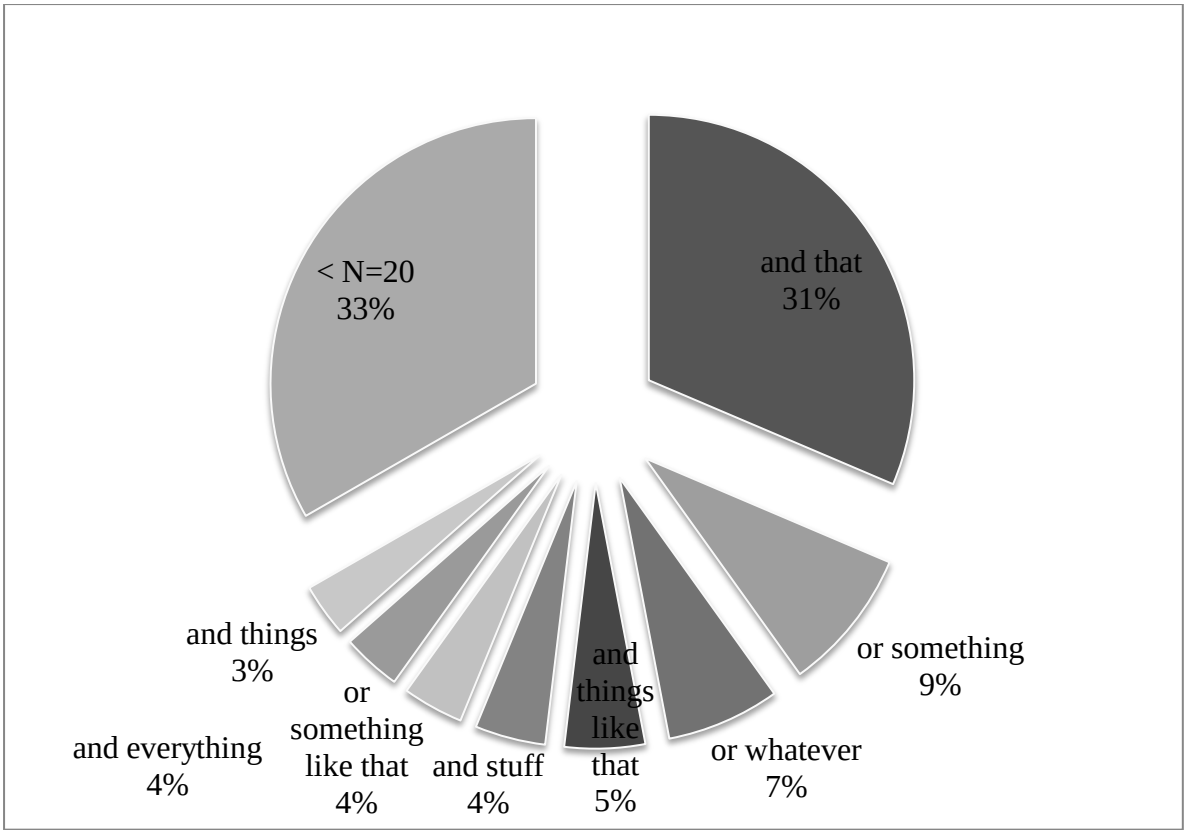

Figure 1. Overall distribution of productive variants

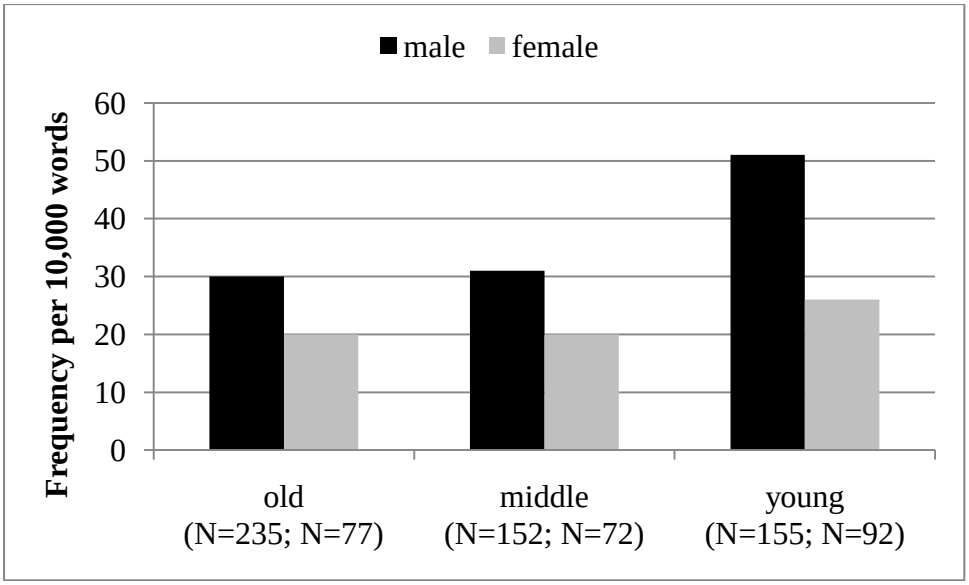

Figure 2. Frequency of GEs per 10,000 words by age and sex