CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/57465/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Tagliamonte, Sali A., Durham, Mercedes and Smith, Jennifer 2014. Grammaticalization at an early stage: future 'be going to' in conservative British dialects. English Language and Linguistics 18 (1) , pp. 75-108. 10.1017/S1360674313000282

Publishers page: http://dx.doi.org/10.1017/S1360674313000282

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Post peer review version.
Sali A. Tagliamonte, Mercedes Durham and Jennifer Smith (2104). Grammaticalization at an early stage: Future be *going to* in conservative British dialects. *English Language and Linguistics* 18.1: 75–108. Cambridge University Press doi:10.1017/S1360674313000282

GRAMMATICALIZATION AT AN EARLY STAGE: FUTURE BE *GOING TO* IN CONSERVATIVE BRITISH DIALECTS[1]

SALI A. TAGLIAMONTE[1], MERCEDES DURHAM[2], JENNIFER SMITH[3]

[1]*University of Toronto* [2]*Cardiff University* [3]*University of Glasgow*

FIRST AUTHOR'S CONTACT DETAILS

Sali A. Tagliamonte
Department of Linguistics
University of Toronto
100 St George Street
Room 4077 Sid Smith Hall
Toronto, Ontario,
Canada M5S 3H1
E-mail: sali.tagliamonte@utoronto.ca

[1] The authors gratefully acknowledge the Economic and Social Research Council of the United Kingdom (Tagliamonte, Smith), the Arts and Humanities Research Board of the United Kingdom (Tagliamonte), and the Carnegie Trust for Universities in Scotland (Durham) for their generous support in conducting this research and writing up the paper. We also thank our ELL reviewers and editors who contributed greatly to this final product, although none but ourselves are responsible for its present content.

ABSTRACT:

The English *go* future, a quintessential example of grammaticalization, has shown layering with *will* since at least 1490. To date, most synchronic evidence for this development comes from dialects where *be going to* represents a sizable proportion of the future temporal reference system. However, in the United Kingdom in the late 20th century there were still dialects where *be going to* was only beginning to make inroads, representing a mere 10-15% of future contexts. These varieties offer an effective view of the early stages of grammatical change.

Statistical analysis of nearly 5000 variable contexts reveals that the use of *be going to* is increasing across generations, but at different rates, depending on location and orientation to mainstream norms. Major patterns of use mirror previous findings: *be going to* is favoured for subordinate clauses. However, other widely reported constraints conditioning *be going to* are radically different across age groups, exposing contrasts between incipient vs. later stages of grammaticalization. In the most conservative dialects *be going to* is strongly correlated with negatives and questions especially in 1st person singular. This suggests that these contexts may have been the 'trigger' environments for redistribution of meaning of the incoming grammatical form (Hopper & Traugott, 1993:85). The fact that strong effects of negatives and questions endure in contemporary urban varieties (Torres-Cacoullos & Walker, 2009) confirms that grammaticalization begins in very specific syntactic contexts, and impacts the system for generations to come. In contrast, other reported constraints - resistance of *be going to* in 1st person singular and extension to inanimates and far future readings - emerge across generations, suggesting they are later developments.

Taken together, these findings demonstrate how synchronic dialects show us incremental steps in the grammaticalization process. Comparative sociolinguistic analysis thus offers insights into which patterns define the point of grammaticalization itself; which derive from systemic processes; which can be attributed to discourse routines and collocations; and how these factors converge in shaping the evolution of grammar.

GRAMMATICALIZATION AT AN EARLY STAGE: FUTURE 'BE GOING TO' IN CONSERVATIVE BRITISH DIALECTS

"The study of grammaticalization can be understood as an attempt to disprove the assumption that changes resulting in grammatical forms are completely random and unpredictable" (Hopper and Traugott 1993: 232)

1. THE FUTURE

A quintessential example of grammaticalization is the so-called *go* future, which not only reveals over-arching cross-linguistic pathways (Givón, 1979, Bybee, Perkins & Pagliuca, 1994, Heine & Kuteva, 2002) but also well defined developmental constraints. In English the *go* future is the construction *be going to* which has been involved in a longitudinal layered relationship with *will*, as in (1), for the past 600 years, perhaps longer.

(1) a. Oh no, I *ain't gonna* do that, he *won't* talk if ee do that. (TIV/a)[2]

b. I think that's where it'*s gonna* be most expensive so that*'ll* be savings plus wages to pay for that, but it*'ll* be fine when we get it all done. (LER/6)

The development of the *go* futures is one of the most widely studied grammaticalization pathways. Research on the *go* future spans different languages (French, Spanish, English, pidgins and creoles, etc.), research paradigms (descriptive, theoretical, empirical) and methodological approaches.[3]

In contemporary English the *go* future is still evolving, so the question is: what is its current position on the diachronic trajectory of change? In the late 20th and early 21st century, the future temporal reference system remains extensively layered, split between *will*

---

[2] The examples come from the data under analysis. The community is identified by a three-letter acronym followed by a slash and then a single digit for speaker.

[3] See e.g. Curme, 1913, Royster & Steadman, 1923/1968, Fries, 1925, Fries, 1927, Luebke, 1929, Ultan, 1972, Harner, 1976, Wekker, 1976, Close, 1977, Anderson, 1979, Harner, 1980, Haegeman, 1981, Ofuani, 1981, Fleischman, 1982, Wales, 1983, Nieuwint, 1986, Bybee & Pagliuca, 1987, Elson, 1988, Myhill, 1988, Nehls, 1988, Arnovick, 1990, Gagnon, 1990, Bybee, Pagliuca & Perkins, 1991, Howe & Strauss, 1991, Grancrov, 1992, Schwegler, 1992, Roberts, 1992-3, Danchev & Kytö, 1994, Berglund, 1997, Mair, 1997, Aceto, 1998, Poplack & Turpin, 1999, Danchev & Kytö, 2002, Tagliamonte, 2002, Torres-Cacoullos & Walker, 2009, Wagner & Sankoff, 2011).

(including the contracted form *'ll* and the negative form *won't*) and *be going to* (and its phonological variants) along with use of the present and present progressive. However, in the vast majority of cases, future meaning is expressed by variants of *will* and *be going to*. According to many commentators these two forms, despite their divergent origins, exhibit "no demonstrable difference" between them (e.g. Hall & Hall, 1970, Palmer, 1974:163, Quirk, Greenbaum, Leech & Svartvik, 1985:218, Danchev & Kytö, 1994:384). Indeed, in the data we will be investigating in this study from a number of dialects across the British Isles (see Section 3), alternation of forms is rampant, often occurring in the same sentence in the same stretch of discourse, as in (1) above and (2).

(2) a. I says, "It'*ll* only be six month." Didn't know it *were gan* be six year. (MPT/r)

b. "How you *gonna* do that?" "Oh," said I, "I*'ll* soon do it." (CMK/m)

c. You*'ll* have to marry Wendy, because otherwise she*'s not gonna be* able to ... afford to live. (HEN/e)

Considerable study has been devoted to the typical evolutionary path of *go* futures and there is general consensus of a common path of development across languages. We will make use of these pervasive pathways of change to inform our hypotheses and interpret our analyses of *be going to* in our data. First, we turn to a discussion of the chronological evolution of the grammaticalized form *be going to* in English.

2. EVOLUTION OF THE FUTURE

The use of *be going to* as a future marker originated as the progressive aspect of the lexical verb *go*, meaning movement towards a goal, which collocated with a preposition followed by a complement.[4] A cursory perspective on what the language was like at earlier points in time can be gleaned from the Helsinki Corpus (1991), a multi-genre diachronic corpus which spans Old, Middle and Early Modern English. A gross extraction of all forms of *be going*

[4] The present data do not permit us to address how *be going to* changed from subcategorizing an NP to a verbal complement.

to/goyng to returns just 48 tokens, almost none of which can be construed as having future temporal reference. This is consistent with Mair (2004), who suggests that the grammaticalized version was rare until the 19th century. In contrast, there are 1838 tokens of *will* and 95 of *'ll.* One of the prerequisites for early grammaticalization is the occurrence of ambiguous bridging contexts, in which "both the old and the new meaning can occur" (Brinton & Traugott, 2005:109); in the Helsinki Corpus, the only reasonable example meeting this criterion is the one in (3)a, which provides the important context of intention in the development of *be going to* (e.g Poplack & Tagliamonte 1999; see also Disney 2009). In most cases, substitution with *will* is not possible, as in (3)b. At this point in time the construction *be going to* is also used to mean "intend". Concomitantly, the verb *intend* is used in contexts that are future-like in meaning, as in (3)c.

(3) a. I will tell it M=r= Isaac. S=r= John Walter is going to be marryed to my Lady Stoel w=ch= will be very happy for him. (Helsinki Corpus, 1695, The Countess of Nottingham)

b. The council sat upon it, and were going to order a search of all the houses about the town. (Helsinki Corpus, *The Popish terror and the impeachment of Danby*)

c. My Cozin Val Pettit is paid his debt. Mr. Trusser's bond and Mr. Dickenson's bond I intend, God willing, to pay tomorrow being Mooneday; if I see Mr. Twiman I intend to pay him what is due to Him; and if Mr. Crux his bond be sent up, I intend to pay that, so that I shall not be troubled with their summons any further. (Helsinki Corpus, 1662, Henry Oxinden)

These facts make it understandable why the earliest date of attestation of an unambiguous use of *be going to* to express future temporal reference has been somewhat contentious; however, we can reasonably assume that *be going to* started being used with future temporal reference sometime during the mid 1400's, based on the examples in (4) and (5) reported in Danchev and Kytö (1994, 2002).

(4) Therefore while this onhappy sowle by the vyctoryse pojmpys of her enmyes *was going to be* broughte into helle for thesynne and onleful lustys of her body ('Therefore, while this unhappy soul by the formidable victories of her enemies was going to be sent to hell for the sin and wicked lusts of her body') (Helsinki Corpus, 1482, *The Revelation to the Monk of Evesham)* (Danchev & Kytö, 1994:69)

(5) And thane come Englissh folk to the seid Merchauntz of the Maryknyght and bad theym beware whome they had lefte yn their Ship sayng that yt was likely be taken And there vppon the seid persones of the ship of Hull *goyng to do* the said wrong/ yaf to oon henry wales Gentilman… (Helsinki Corpus, 1438, *Chancery English*) (Danchev & Kytö, 2002)

It has been suggested that the early meaning of physical movement towards a goal (e.g. *be going to be married)*, gradually receded in favor of a more general sense of prediction (Royster & Steadman, 1923/1968:402) and connotations of intention, purpose and determination came to the fore. Table 1 summarizes these developments in relation to other forms expressing future meaning (*will, shall*), highlighting the interaction of forms across time.

Old English	16th century	17th century	Modern English

shall: present obligation *will:* •volition •willingness	*shall:* simple future *will:* modal future *be going to:* actual motion	*shall:* 1st p. future *will:* •2nd+3rd p. future • prescriptive rules of usage *be going to:* •immediate or impending •correlated w/ motion verbs	*shall:* •restricted •formulaic •infrequent *will:* simple future *be going to:* ?? future

Table 1:
Historical Perspective on developments of future temporal reference forms.

By the 17th century *be going to* became more frequent and is reported with a wide array of lexical verbs while still retaining strong associations with its literal meaning of "intention" and "movement" (Danchev & Kytö, 1994). Eventually *be going to* started occurring with inanimate subjects and stative verbs. The prevailing story in the literature suggests that *be going to* has been gaining ground ever since (e.g. Mair, 1997, Poplack & Tagliamonte, 1999, Tagliamonte, 2002). Recent books reporting frequency data on *shall/will* in present day English converge in reporting: 1) an increase in *be going to* over time; 2) a heightened rate in North American English and 3) differential usage by register with press reporting and spoken data in the lead Leech et al. 2004:108,271, 294 etc.; Mair 2006:95-100. This is supported by a Google n-grams search for *am be going to/is be going to/are be going to* in British and American fiction where there is an upwards swing across both varieties over the period 1840-2000.[5]

Contemporary reports suggest that *shall*, as in (6)a, has mostly receded to formulaic and legal uses (Williams, 2013). Present tense and present progressive constructions are infrequent and highly circumscribed, as in (b-c) (e.g. Poplack & Tagliamonte, 1999, Torres-Cacoullos & Walker, 2009). The rise of *be going to* appears to be assured.

(6) a. I *shall* have to put the kettle on in a minute. (TIV/ a)

[5] http://ngrams.googlelabs.com/ accessed 9-5-11, 8:19am

b. Aye she's at Workington but she *goes* to Carlisle in September (MPT/ [a])

c. They*'re spending* Christmas and New-Year with her (CMK/ F)

It is still not entirely clear how or when *be going to* evolved to express pure prediction. This is undoubtedly due to the fact that the meanings of intention and motion are so difficult to distinguish (Harada, 1958, Pérez, 1990). Indeed, the early examples in (4) and (5) above have elements of movement, intention and proximity in the future. Thus while meaning change in grammaticalization often proceeds along clines (Heine & Kuteva, 2002), this suggests that some of the most grammaticalized meanings for *be going to* were in place from the very beginning of its future uses long before the form became frequent in the mid-17th century (Royster & Steadman, 1923/1968, Fries, 1940, Pérez, 1990, Hopper & Traugott, 1993, Danchev & Kytö, 1994)[6]. Indeed, recent research on Quebec City and Montreal English in Canada in the early 2000s (Torres-Cacoullos & Walker, 2009) suggests that many of the earliest meanings remain entrenched. This brings into question the relative timing of different stages of grammaticalization and the role that frequency of use plays in the process.

3. DATA AND METHODS

The goal of this paper is to add to the available information on grammatical development by measuring the pathway of change of *be going to* in a compendium of vernacular speech data from research projects in the United Kingdom at the turn of the 21st century.[7] To study grammatical change in synchronic data we first need the perspective of time. In this case we will use two different proxies for time depth. On the assumption that language change progresses across space (e.g. Trudgill's (1974) Gravity Model), we will make use of a range

[6] But see Mair's (2004: 126) suggestion that we are "still far from a precise understanding of the role of frequency in grammaticalisation". His analysis of *be going to* in the OED shows that while the form is said to have grammaticised as far back as the 1600s, its frequency in use does not increase until around the 19th century. Mair (2004) distinguishes two types of grammaticalization. The first is "a dynamic type which involves diachronic change and will result in drastic shifts in discourse frequencies of the constructions concerned" and includes *be going to*. The second "a static type, is best described as the occasional grammatical use of lexical material" (ibid: 123). In other words, some types of grammaticalization may be linked to frequency increases, but others may not (see also Hoffmann 2005).

[7] To our knowledge this collection represents the largest body of spoken vernacular English to be subject to analysis for future temporal reference.

of dialects across the UK. On the assumption that language change is represented among people of different ages, we will use the generations of the communities in which the dialects are spoken. Second, we need an analytic method designed to examine complex patterns of language use in conversational data. In this case, we will exploit both the frequency of forms as well as the underlying constraints on their competition. This will expose not only the progress of grammaticalization, but also the mechanisms of change, as represented in the shifting weights and strength of different influences in the process. Layering – the co-existence of multiple forms "within a broad functional domain" (Hopper, 1991:22) – between the main expressions of future temporal reference (*be going to* and *will/'ll*) will be the foundation of our analysis. We will begin with systematic exploratory study of the data, using comparison of marginals and cross-tabulation to arrive at a detailed understanding of the patterns inherent in the data set and conditional inference trees to uncover how the most important predictors work together (e.g. Tagliamonte & Baayen, 2012:163-164) and finally a series of mixed effects generalized linear models to evaluate the statistical significance, patterns and interactions across competing contextual constraints influencing the use of one variant or the other (e.g. Baayen, 2008). Interpretation of the results of the statistical models compared across communities and across generations (see, e.g. Tagliamonte, 2002) will be used to help us situate *be going to* as it grammaticalizes.

A further consideration will be the extent to which grammaticalization pathways are parallel from one major variety of English to another. On the one hand, grammaticalization is thought to be a phenomenon that has broad typological tendencies. On the other hand, local idiosyncracies may lead to divergent tendencies. It would not be surprising, of course, for varieties to 'go their own way' since grammaticalization is well known to be influenced by social factors (Hopper & Traugott, 1993). Indeed, a single constraint may have varying tendencies from one dialect to another. Thus, in a cross-variety study of the preterit of *be,* Tagliamonte (2011:149-151) discovered that the effect of negation differed across communities. Some varieties favoured *was* for negatives, some varieties favoured *was* for affirmatives. In other cases there was no difference between the two. Such possibilities will inform our interpretation of trends in one locality compared to another.

Crucially for our purposes, *be going to* is said to be more frequent in North America than the United Kingdom (Wekker, 1976, Berglund, 1997, Tagliamonte, 2002, Szmrecsanyi,

2003). This 'décollage' between two major varieties of English offers the analyst a unique opportunity to examine how grammaticalization proceeds since it may be that these two major varieties are at two different points in the trajectory of change.

3.1 Communities

Our analysis draws on data from ten communities across the United Kingdom – from the Shetland Islands in the North to Devon in the South, as in the Map in Figure 1.[8]

Figure 1: *Communities studied*

This data was collected for a number of different projects between 1997-2009 (Tagliamonte, 1996-1998, Tagliamonte, 1999-2001, Tagliamonte, 2000-2001, Tagliamonte, 2001-2003, Smith, 2007-2009) as in Table 3. The communities range from small villages to cities, from the Shetland Islands off the north coast of Scotland to the southern counties of England, from mainstream to isolated. All the data come from sociolinguistic interviews, which are

[8] Each community is represented by markers which have the community initials.

essentially conversations focused on oral history and personal narrative (see Labov, 1972). In four of the communities (York, Wheatley Hill, Buckie and Lerwick) we collected interviews from speakers ranging from young adults to the oldest residents, which will allow us to also study change in apparent time. For the others (Cumnock, Portavogie, Cullybackey, Maryport, Tiverton, Henfield) only the oldest generation was surveyed[9] (see Tagliamonte, Smith & Lawrence, 2005:87-117, Tagliamonte, 2013b:27-45 for a fuller description of these communities).

Community [abbreviation]	Male	Female	Total Speakers	Total Words
Scotland				
Lerwick [LER]	15	15	30	347,785
Buckie [BCK]	20	20	40	290,000
Cumnock [CMK]	18	23	41	349,428
Northern Ireland				
Cullybackey [CLB]	12	5	17	198,086
Portavogie [PVG]	7	2	9	92,803
North of England				
Maryport [MPT]	20	23	43	401,376
Wheatley Hill [WHL]	12	11	33	253,497
York [YRK]	39	52	91	1.2 million
South of England				
Henfield [HEN]	4	4	8	125,000
Tiverton [TIV]	7	2	9	96,472
Total	154	157	311	3.35 million

Table 2:
Sample Design

3.2 Circumscribing the variable context

The future temporal reference system is often highly layered and there is considerable variation in terms of where (and how often) different future variants can occur and which of

[9] The two Northern Ireland communities, Portavogie (PVG) and Culleybackey (CLB), with total Ns of only 26 and 301 respectively across all future tense markers, were combined for the logistic regression analyses.

these are included across studies. In this study we followed the data coding practices of earlier work (Poplack & Tagliamonte, 1999, Tagliamonte, 2002, Torres-Cacoullos & Walker, 2009) and included for analysis only those expressions that were clearly temporal and made reference to future time. While various tense/aspect constructions can encode future temporal reference in English, the simple present and the present progressive tend to be highly circumscribed. The present tense for example is mostly restricted to temporal clauses where there is an accompanying time adverb (Mossé, 1952, Biber, Johansson, Leech, Conrad & Finegan, 1999:454-455), or with near future associations (Torres-Cacoullos & Walker, 2009:330), as in (7). Moreover, contemporary quantitative investigations have found that these mark future temporal reference relatively infrequently (Poplack & Tagliamonte, 1999:326, 5-16%, Table 2) (Torres-Caccoulos & Walker 2009:328, 3%, Table 1).

(7) She *goes* to Carlisle in September (MPT/ª)

Another anomaly in the contemporary system, as mentioned earlier is *shall*, which is often reported to be restricted to formal registers and formulaic utterances in modern English (Williams, 2013). Furthermore, it is rare to non-existent in Ireland and Scotland, (Robertson & Graham, 1952/1991, Crystal, 1986). Indeed, we found only 43 tokens of *shall* and of these 51% (N = 22) come from York, the only urban centre; they are used mostly by women and in formulaic questions, as in (8).

(8) a. What shall I do? (YRK/Z)

b. *Shall* I put this back on? (YRK/R)

The periphrastic expression *about to*, as in (9), was very rare (N=4):

(9) a. I was just *about to* say he must've been clean-living- but no- he smoked. (MPT/8)

b. They were *about to* settle in Australia but the ship had got- it was shipwrecked. (MPT/10)

All these forms were excluded from the variable context in order to focus in on the robust variability between variants of *will* and *be going to.*

Further constructions were found to be invariant and these were also excluded. For example, tag questions always agree with the verb in the main clause (10).

(10) Oh, you'll spoil my chances, *will* you? (BCK/024)

Future-in-the-past contexts, as in (11), represented 5% of the total number of tokens (N=278). Of these 98% are *be going to*, making them near-categorical in our data.

(11) a. It was like a bitty more than we thought it *was gan* to be. (BCK/j)

b. I was terrified he *was gonna* tell the whole class that I'd cried. (MPT/%)

These contexts are substantially different from standard future temporal reference in terms of frequency of forms (predominance of *be going to*) and patterns of use (antithetic constraints) (Poplack & Tagliamonte, 1999:334, Torres-Cacoullos & Walker, 2009:327) and therefore we have excluded them here as well.

The resultant data comprises 4764 tokens of *will* and *be going to* from 183 individuals born between 1906 and 1989 across 10 communities.[10]

4. RESULTS

As a first step in the long process of understanding the trends and patterns in the data, we begin with a comparison of marginals of the linguistic and social predictors and their interactions. The proportion of *be going to* out of the total N of *will* and *be going to* combined, according to community and three broad age groupings (≥ 31, 36-65, and 66+), is displayed in Table 3.

[10] Compare the figures in some earlier studies: 3337 from Quebec City English (Torres-Cacoullos & Walker, 2009), 2561 from Toronto English (Tagliamonte & D'Arcy, 2009); 1330 from York English (Tagliamonte, 2002).

	Age Cohort					
Community	≥ 31		31-65		66+	
	%	N	%	N	%	N
N. Ireland					**9**	312
Buckie	**21**	549	**17**	298	**12**	314
Cumnock					**12**	177
Lerwick	**25**	268	**24**	284	**19**	122
Henfield					**19**	151
Maryport					**21**	185
York	**41**	433	**24**	300	**21**	256
Wheatley Hill	**26**	284	**20**	492	**23**	178
Tiverton			30	27	**22**	96
Overall	**28**	1534	**21**	1466	**16**	1764
	Overall **22%**			Total N: 4764		

Table 3:
Rates of be going to *by community/age cohort*

The gaps in our coverage of the population in these communities (the greyed areas in Table 3) are immediately apparent. In Northern Ireland, Cumnock, Henfield, Maryport and Tiverton, only the oldest generation was sampled. As noted earlier, this is due to the distinct research questions of the four independent research projects from which these materials have been drawn. These gaps notwithstanding, Table 3 provides a striking perspective on how different the proportion of *be going to* can be depending on the age of the individual and community (compare York residents ≥31 at 41%, indicated by the bold borders compared to the oldest people in York, 21% and Northern Ireland, 9%).[11] These facts are crucial and demonstrate that no study of *be going to* should be undertaken without contextualizing the materials in terms of the region from which they come, the age of the speakers in the sample and the date of collection. A compounded, albeit separate issue, is the problem of low

[11] The apparent time distribution in Table 3 shows a situation of surprisingly modest development in Wheatley Hill and Lerwick, but more robust change in Buckie and York. We take this to be an indication that grammaticalization of *be going to* is moving at a different pace in different localities. This calls for further research on this feature in these communities in the near future, including in-depth cross-community comparisons of change in apparent time.

numbers. The nature of the sample design will impact how far the data can be partitioned before the diminishing token counts forestall further breakdown of the data. For example, the Northern Ireland data (N= 312) already represents a combined sample from Portavogie and Culleybackey since the number of future temporal reference contexts in Portavogie was so limited (N=26). All these issues directly affect our approach to these materials.

It is, of course, critical to analyze *be going to* in data sets that are in fact comparable. Figure 2 shows the proportion of use of *be going to* among the age group for which we have cross-dialectal representation — speakers aged 66 or older. The total N for this analysis comprises 1764 tokens. Figure 2 also includes the proportion of *be going to* in Toronto (c. 2003-4) (Tagliamonte & D'Arcy, 2009, Tagliamonte, 2012) for comparison.[12]

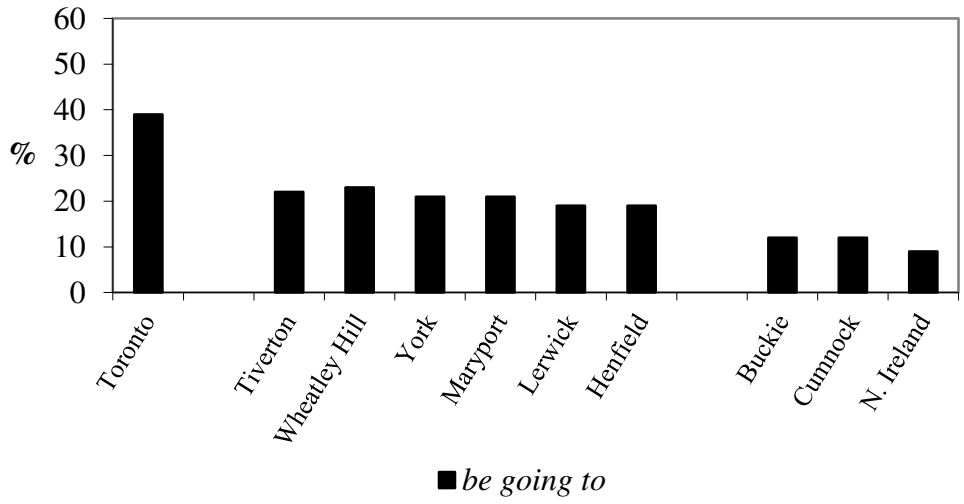

Figure 2: *Overall proportion of* be going to *by community, speakers over 66.*

[12] Other communities in North American have been studied quantitatively with comparable rates of *be going to,* e.g. Quebec City and Montreal (c. early 2000s), (Torres-Cacoullos & Walker, 2009) (c. early 2000's), Ottawa (c. 1990s) (Poplack & Tagliamonte, 1999) (c. 1990's), and American English (from the Santa Barbara Corpus of Spoken American English, c. late 1990's) (Szmrecsanyi, 2003). These corpora suggest a much higher use of *be going to* than is found in the conservative dialect data; however, due to the fact that they included individuals of all ages this can only be supposition without being able to probe these materials more accountably.

The data in Figure 2 reveals that the frequency of *be going to* amongst the oldest generation in the United Kingdom communities is indeed comparatively moderate (each community 23% or under) in comparison to North American varieties (over 35%). Further, in the United Kingdom, the communities can be distinguished according to the frequency of *be going to*. It is lowest of all, 9%, in the villages of Northern Ireland (Portavogie and Cullybackey), and small towns of Scotland (Buckie and Cumnock). In contrast, it is 20% or more in Tiverton, Wheatley Hill, York, Maryport, Lerwick and Henfield. Thus, it seems plausible to assume we have caught the encroachment of *be going to* at an incipient stage of its development in these communities.[13] At the same time the divergent frequencies correlated with speaker age and community add a high degree of complexity to this data structure. How can we approach the analysis so as to discover the process underlying the variation?

The overall frequency of the incoming form (*be going to*) is in line with previous analyses of other features in these varieties. Buckie, Portavogie and Culleybackey are consistently found to be the most conservative across a number of linguistic forms (Tagliamonte & Smith, 2005, 2006, Tagliamonte et al., 2005).[14] The literature informs us that grammaticalization is typically identified by an increase in frequency of the grammaticalizing form (e.g. Bybee, 2003). Indeed Brinton and Traugott (2005:209) suggest that "token frequency can be used to hypothesize the historical time-depth of a particular grammatical morpheme – the more frequent an item the more grammaticalized it is." Thus, based on our previous results for these varieties, coupled with the frequencies of use in Figure 2, one way to approach an analysis of these communities is to arrange them in terms of the relative frequency of *be going to*[15]. We can also make profitable use of Nevalainen and Raumolin-Brunberg's

---

[13] Although the *be going to* construction has been attested for future temporal reference as far back as the 15th century (cf. examples (4) and (5) above), these reports are based on written materials that come from mainstream areas. One of the goals of this investigation is to show that this grammatical change has proceeded at very different rates across dialects in the United Kingdom.

[14] Lerwick has not previously been compared to these other varieties.

[15] One reviewer suggests using a measure of phonetic assimilation (*gonna*) to distinguish the different communities. However in at least two of the communities, BCK and LER,

assessment of how language change proceeds (Nevalainen & Raumolin-Brunberg, 1996:213-255). For them, an incoming form is 'incipient' when it is used at a rate of 15% or below, while it is 'new and vigorous' when the incoming form is between 15 and 36%. Using Nevalainen and Raumolin-Brunberg's division suggests a means to partition the communities based on these frequency distinctions as a proxy for viewing the stages of grammaticalization of *be going to*. In some communities, it has just begun to enter the future temporal reference system, while in others it has moved further along the path of grammaticalization. By comparing the communities in this way, we may be able catch *be going to* as it transits from one level of frequency to another, and potentially from one stage of development to another. In Canadian varieties considered previously "early stage" communities (Poplack & Tagliamonte, 1999) had rates of *be going to* well within the "vigorous" levels defined by Nevalainen & Raumolin-Brunberg (1996) which further supports our contention that we have caught the grammaticalization of *be going to* on the upswing of the S curve in these conservative dialect materials.

4.1 Constraints in apparent time

Further support for this approach to the data can be gleaned from investigating the apparent time dimension of the data. Figure 3 shows use of *be going to* across the four communities where we have generational data (Wheatley Hill, York, Lerwick and Buckie). This view of the data provides a visualization of the frequencies of *be going to* by the generational cohort and community (see Table 3 for the Ns and %s).

---

phonetic assimilation of the Scots *gan* + *to* is not possible, thus this cannot be used as a diagnostic.

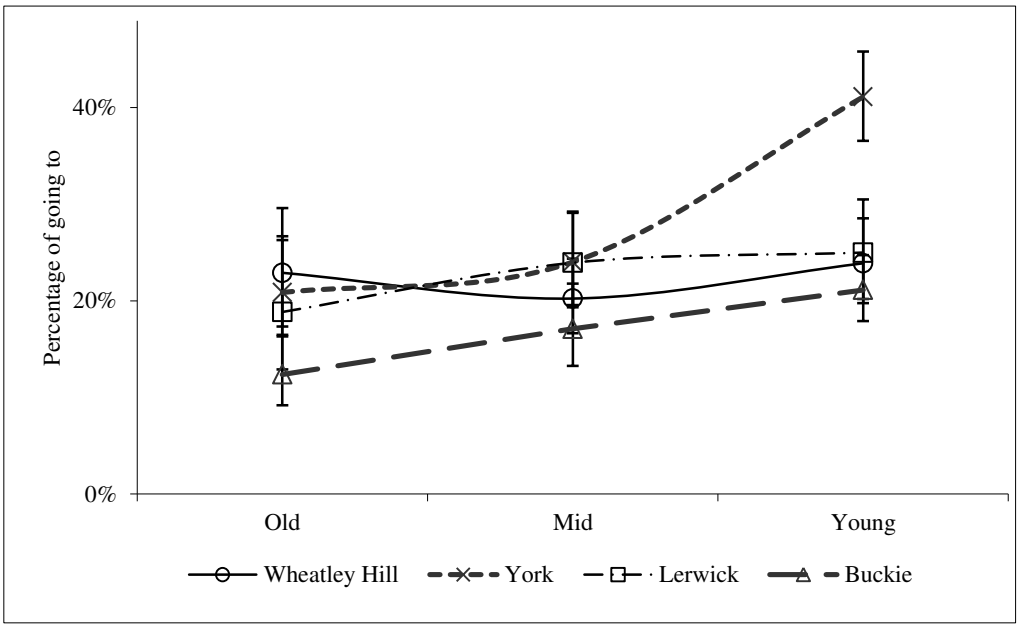

Figure 3: *Apparent time perspective on the use of 'be going to' in four communities.*

The future temporal reference system in Lerwick and Wheatley Hill is either stable or changing very slowly.[16] There is a modest increase in the use of *be going to* in Buckie but a distinct upswing in the youngest speakers in York (Tagliamonte, 2002). The difference between the age groups is statistically significant in both communities, but obviously stronger in York. Note too that the error bars confirm that the York younger speakers are evolving well beyond the rest of the population, suggesting that they are the vanguard of the change.

Complex interactions and vastly different cell sizes such as in the data structure under investigation are known to be difficult to capture adequately even with a mixed-effects logistic linear model. In order to determine which communities and age cohorts to distinguish so as to model appropriate stages of development we sought statistical corroboration by subjecting the data to a conditional inference tree (Strobl, Malley & Tutz, 2009), as in Figure 4.[17] The value of this method is that it can identify the complex

[16] Torres-Cacoullos and Walker (2009) report an increase in *be going to* for the youngest speakers in Quebec City and Montreal (c. 2002-2003) and "slightly more in Montreal"; however, they do not provide any data or figures.

[17] Z = Shetland, Lerwick (L), P = Maryport (M), X = Sussex, Henfield (H), D = Devon, Tiverton (T), A = Cumnock (C), I = Northern Ireland (Culleybackey (Cb) and Portavogie (P).

interactions characterizing a data set (for a practical example see Tagliamonte & Baayen, 2012).

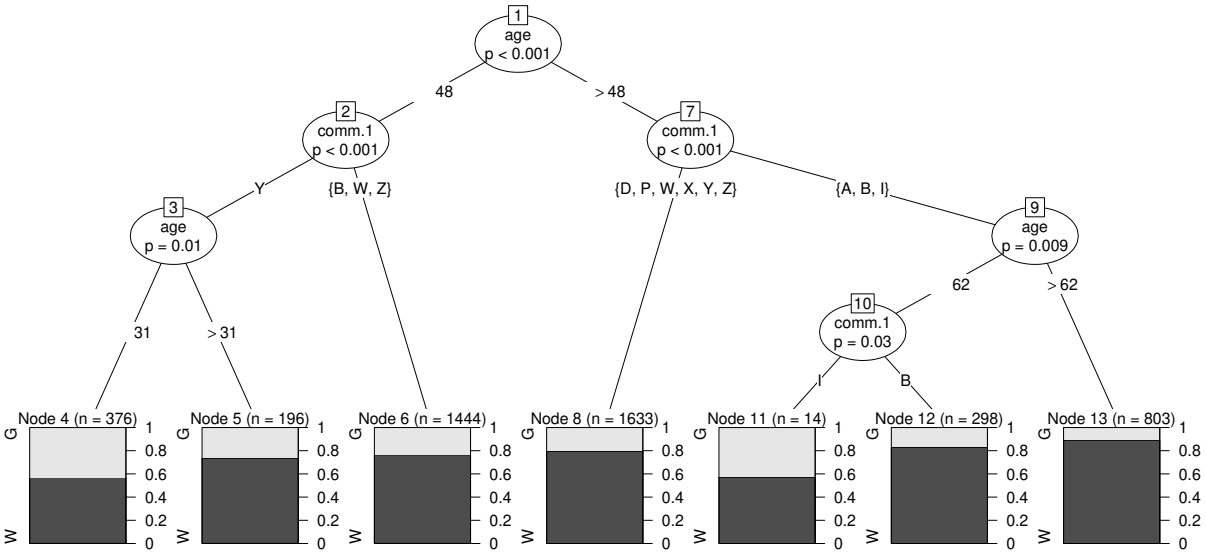

Figure 4: *Conditional Inference Tree – community and speaker age*

In this case, the conditional inference tree uncovers the relevant important divisions in the data.[18] It shows that are five major groupings according to community and age cohort, which corresponds to varying frequencies of *be going to*, as illustrated in Table 4.[19]

Community	Age	Frequency	Stage
York	≤ 30	43%	vigorous
Buckie, Wheatley Hill, Lerwick, York	>48 31-48	28%	active

[18] In some cases we have grouped categories together that patterned similarly in the subsequent regression analyses. In addition, the Northern Ireland data with very small cells (N=14) was grouped with the elderly individuals in the most conservative group (Cumnock and Buckie).

[19] Diverging somewhat from Nevalainen and Raumolin-Brunberg's labels, we will refer to the frequency levels of 21% and 28% as "active" and the highest frequency situation among the York young speakers as "vigorous".

Tiverton, Maryport, Wheatley Hill, Henfield, York, Lerwick	> 48	21%	active
Cumnock, Buckie, Ireland	<62	13%	incipient

Table 4:
Community/Age partitions for statistical modelling

In the analyses that follow, we will partition the data for statistical modelling as follows: 1) communities and age cohort where *be going to* is incipient (<15%); 2) communities and age cohorts where it is active (21-28%)[20] and 3) the one community (York) where the youngest age cohort stands apart from the rest in vigorous use of *be going to* (43%).[21] We interpret the first as providing insight into varying stages of the entry of *be going to* as a future marker into the grammar and the latter as providing insight into how it has spread. Just as the development of *be going to* did not start at the same time in all communities, neither can we assume that it will develop at the same speed in all communities or even advance in a regular and consistent way. Wagner & Sankoff (2011), for example, show that the development of the periphrastic future in French was arrested along its trajectory of change and Tagliamonte and Smith (2006) have documented a reversal in the development of modal *have got to*. Thus in addition to probing the rates of change by locality and generation, we will pay particular attention to how the well-known constraints on the grammaticalization of *be going to* as a future marker in English are reflected in these generational cohorts in the dialects.

[20] Analysis of other features in Shetland revealed stark differences between the young speakers in use of traditional dialect features, with half using the older, dialectal variants and half using the newer, more Standard variants (Smith & Durham, 2011, 2012). We do not, however, find this intra-group variation for *be going to,* likely due to the fact that there are no dialect-specific variants.

[21] Several analyses (not shown) revealed that there was only marginal significance in the difference between the two "active" groupings, so the 31-48 year olds have been collapsed as "active" in the analyses that follow.

4.2 Testing for grammaticalization

In what follows we will review the main predictors reported for the development of *be going to* (Bybee et al., 1994) using the research results from earlier quantitative analyses as a baseline (Poplack & Tagliamonte, 1999, Torres-Cacoullos & Walker, 2009), as summarized in Table 5. The table displays a composite view of overall developments attested for each predictor: sentence type, clause type, proximity, animacy and grammatical person.[22] Poplack and Tagliamonte's (1999) analysis was based on a range of communities arrayed from conservative to mainstream. Canadian enclaves in Nova Scotia were argued to be at an early stage of development, with data from the Canadian capital city of Ottawa showing a later stage[23]. Torres-Cacoullos and Walker's (2009) study was based on data from Quebec City and Montreal, which they argued represented a late stage in the grammaticalization of *be going to*. These studies, their results and their varying time and place enable us to contextualize and interpret what situation we find in the United Kingdom communities where *be going to* is either incipient or vigorous.

	EARLY STAGE	LATE STAGE	CHANGE
Sentence Type	Interrogatives and negatives highly favour; interrogative > negative	Interrogatives and negatives highly favour; interrogative > negative	No change

[22] Hilpert (2008) utilizes a Construction Grammar approach to analyse statistical co-occurrence patterns between future constructions and lexical items across a number of Germanic languages with the aim of uncovering semantic change in this area of the grammar. His results from three time periods in the history of English - 1710-1780, 1780-1850, 1850-1920 – show that *be going to* collocates with telic and dynamic verbs in the first two time periods. In contrast, in the third time period, it collocates with the most frequently occurring verbs more generally, including the stative verbs *be* and *have* suggesting that the meaning of *be going to* has generalized, i.e. become more grammaticalized. We tested for a range of collocations in our data, but none were significant.

[23] Note that Poplack and Tagliamonte 1999 included future-in-the-past contexts in their analyses so these comparisons may not be precise.

Clause type	Subordinate clauses favour	Subordinate clauses favour	Expansion into main clauses
Proximity	Non-immediate favours over proximate	Proximate favours	Proximity effect develops
Animacy	Animates favour	Animate = Inanimate	Expansion into inanimate contexts
Grammatical person	Non-first favours	Non-first favours	Strengthening of $2^{nd}/3^{rd}$ person for *going to* (1^{st} person favours *'ll*)

Table 5:
Predictions for stages of grammaticalization of be going to

4.3 Sentence type

One of the clearest findings in contemporary studies is the contrast between negative and affirmative, with *be going to* favoured in negatives, as in (12) (Berglund, 1997, Szmrecsanyi, 2003, Torres-Cacoullos & Walker, 2009).[24] Interrogatives also favour *be going to* where these have been studied independently (Torres-Cacoullos & Walker, 2009).

(12) a. They*'re* <u>*not*</u> *be going to* do that. (LER/h)

b. I*'m* <u>*no*</u> *gonna get* that phone number again. (CMK/o)

c. He *was*<u>*nae*</u> *going to* give me it. (CLB/f)

d. He*'ll not* better. No, he*'ll* <u>*not*</u>. (CLB/e)

According to Szmrecsanyi (2003) however, this difference is geographically circumscribed. American English prefers *be going to* in negatives, as in (12a-c), but British English prefers *will/'ll*, as in (12d). Why would this be the case? It may be that as grammaticalization proceeds there is a trend towards more use of *be going to* with negatives or it could be that

[24] The effect of negation on the choice of a *go* future is also widely reported for varieties of French (see Wagner & Sankoff, 2011 and references therein).

these two major varieties of English are simply divergent in their grammaticalization pathway with respect to this predictor. Table 4 shows that it may be more complicated than this. While later stage communities in North America disfavour *be going to* with negatives less than early stages, *be going to* is still not a favoured form. Interrogative constructions, as in (13), tend to be very infrequent compared to other sentence types; however, in studies of linguistic change they are often in the lead. There was a strong effect of interrogatives in Quebec City where *be going to* was found at a rate of (yes/no questions 71%), (WH-questions 92%) (Torres-Cacoullos & Walker, 2009:338)[25]. Re-analysis of the data presented in Poplack & Tagliamonte (1999) yielded a corroborating result: interrogatives consistently and strongly favour *be going to* across all the varieties studied.

(13) a. What *are* we *gonna* do when we get there anyway? (YRK/_)

b. *Is* she *gan* to send it up to you? (BCK/g)

Given these results, Szmrecsanyi's claims and the hypothesised development of *be going to* toward increasing use in negatives and interrogatives, we might expect them to disfavour *be going to* across all the United Kingdom communities. We might also expect a strengthening of this effect across the generations as presumably is the case in the North American situation.

4.4 Clause type

A consistent finding for all studies of *be going to* is that it is favoured in subordinate clauses, as in (14), while main clauses disfavour it (Royster & Steadman, 1923/1968, Poplack & Tagliamonte, 1999, Tagliamonte, 2002, Torres-Cacoullos & Walker, 2009).

(14) a. He's hoping Brian*'s gonna* sign. (MPT/%)

b. If I*'m gonna* write, it's gonna be a story, (MPT/%)

[25] Interrogatives were categorized as yes/no questions (29% N=56) and *wh*-questions (8% N=100).

This may be tied to the fact that subordinate clauses are thought to be the original syntactic location of future readings of *be going to.* Indeed, the two early examples (4) and (5) above, cited by Danchev and Kytö, are both in subordinate clauses.

The conservative communities studied by Poplack and Tagliamonte (1999) mirror the effect displayed in Table 4, i.e. the more conservative communities have a much lower frequency of *be going to* and younger people have a higher frequency. In mainstream communities the difference between main and subordinate clauses had become attenuated, presumably as *be going to* spread to a wider range of clause types. Therefore we predict that we will also find this effect across our United Kingdom communities.[26]

4.5 Proximity

Near future has often been claimed to favour *be going to*, as in (15), while the far future favours *will*, as in (15).

(15) a. We're *gonna* sing a hymn this morning. (PVG/%)

b. You*'re gonna* get wet tonight, aren't you! (YRK/™)

c. I*'m going to* make a cup of tea. (CMK/v)

(16) a. I'*ll* do that when I retire (LER/a)

b. In old age we'*ll* sell it (YRK/a)

c. In ten or fifteen years there'*ll* be nae dialect. (PVG/d)

This effect has been reported for certain contemporary (not relic) North American dialects (Poplack & Tagliamonte, 1999, Torres-Cacoullos & Walker, 2009) with the suggestion that it develops as *be going to* grammaticalizes. Yet other varieties that have been studied show an

[26] Different types of subordinate clauses may be differentiated (Torres-Cacoullos & Walker, 2009). Although we initially coded separately for various types, we present only a two-way division between main and subordinate clauses in our analyses since this was the key difference in our data.

early correlation of *go* futures with proximity, e.g. 18th century English (Roy, 2007) and 19th century Brazilian Portuguese (Poplack & Malvar, 2007). Indeed, considering the effect of proximity for North American early vs. late stage situations (i.e. conservative vs. mainstream communities), it becomes evident that the effect of proximity varies. This may be tied to methodological issues where studies have categorized the contexts differently. In order to distinguish these contexts without temporal grounding, we created a three-way division for proximity: near future[27], far future and no reference. It will be particularly informative to see how this effect operates in communities where *be going to* has only just begun to infiltrate the grammar.[28]

4.6 Animacy

When *be going to* started down the path of grammaticalization it is said to have been correlated with human subjects capable of movement, linked to its original meaning of movement towards a physical goal, as in (17)a-b. Thereafter it extended to non-human and inanimate subjects, as in (17)c-d. The underlying mechanism is the metaphorical spread from movement, (17)a, to intention, (17)b, to prediction (17)c-d.

(17) a. Angus *is going to* to bigg a shed. (LER/c)

b. I*'m gan to* go on a diet, ken. (BCK/w)

c. Your two bones *are gan to* go thegither and grind. (BCK/5)

d. I'm hoping that this weather*'s gonna* ease up a bit. (LER/6)

This was substantiated in the North American varieties. As indicated in Table 5, animates and inanimates are parallel in early stage situations, but in the later stage situations *be going to* expands to inanimates and extends to the point of these favouring *be going to* over

[27] For this analysis, near future was defined as those cases where the reference time to the future was imminent, within the next hour or so.

[28] In contrast to English, temporal reference tends not to influence the choice of the *go* future in contemporary French (e.g. Poplack & Turpin, 1999, Wagner & Sankoff, 2011).

animates. For the United Kingdom data, which we have caught at an even earlier stage of development (i.e. incipient, active and vigorous), there is a substantially lower frequency of *be going to* than in North America (see Figure 2). We might therefore expect animate subjects to actually favour *be going to*. Alternatively, because the United Kingdom communities are distinguished by frequency, it could be the case that the two situations will reveal an early developmental shift from animates favouring *be going to* to extension of use with inanimates.

4.7 Grammatical Person

Many studies report that *be going to* is favoured for 2nd and 3rd person subjects. (Wekker, 1976, Poplack & Tagliamonte, 1999, Tagliamonte, 2002, Torres-Cacoullos & Walker, 2009), as in (18). Researchers have argued that this is due, at least in part, to the fact that first person subjects are more likely to show volition than other grammatical persons and this is "a reading said to be associated with *will"* (Tagliamonte, 2002:750).

(18) a. He*'s going to* be a doctor. (YRK/D)

b. They*'re gan to* bide on for another week. (BCK/t)

c. You*'re gointa* get all sorts. (YRK/t)

This effect is reported in North American varieties at both early and later stages, but the effect of grammatical person strengthens from conservative to mainstream varieties. We predict that we may find a similar contrast in our United Kingdom data.

In contrast to earlier studies, in the analyses that follow grammatical person and animacy have been combined into a single predictor categorized as follows: 1st person subjects, 2nd/3rd person animates and 2nd/3rd person inanimates. This is to avoid interaction between grammatical person and animacy since first person subjects are always animate.

Given these predictions for the grammaticalization of *be going to* and the findings that arise from studies that have tested them in spoken language data across varying types of

communities and speaker generations representing incipient, active and vigorous stages of development, we are now in a position to turn to the analyses of the United Kingdom dialects.

5. STATISTICAL MODELLING

We will test the predictors of grammaticalization of *be going to* by using generalized mixed-effect models, (lme4) in R (Team, 2007). First, we will fit one regression model over all the communities/age cohorts to establish the main trends and interactions of the internal predictors. Second, we will probe a particularly intriguing interaction in the data — the intersection of sentence type and animacy/grammatical person. Third, we will fit comparative mixed-effects models for each community/age cohort in order to compare and contrast the main predictors at each 'stage of development' of *be going to*. This type of analysis will enable us to assess the statistical influence of the predictors over and above the effect of the individuals in the sample as well as the interactions of the predictors with each other. In interpreting the results we will make use of several types of evidence from the ensuing tables. The INTERCEPT indicates the overall tendency of the dependent variable (in this case *be going to*) to surface out of all the other potential forms (i.e. *will, 'll, won't*). As well as considering the dependent variable, the intercept is calculated incorporating a reference level for each predictor. We have set the reference level for each predictor as follows: PROXIMITY, far; ANIMACY/GRAMMATICAL PERSON, first; SENTENCE TYPE, declarative; CLAUSE TYPE, main. In the overall model, we set the reference value for COMMUNITY as York given that this is the context in which the frequency of *be going to* is greatest and distinguished from all other community/age cohorts. Each of the predictors that have been considered in the analysis is listed with the results for each level shown in estimated coefficients in log odds using the reference level as the base. A positive value is a favouring effect of *be going to* to occur in that context. The model also evaluates the relative influence of predictors, as indicated by the stars. Three stars correspond to a p value of 0.001, two stars 0.01, one star 0.05 and a small dot is 0.1. The standard metric for complex data structures with social and linguistic predictors is 0.05. Constraint ranking of factors is

inferred from the difference between the reference level at 0 and the values of the estimated coefficients.

5.1 Overall mixed-effects model — all communities/age cohorts

Table 6 presents a complex generalized mixed-effects model of *be going to* in the data structure as a whole (N= 4764) with the communities/age as an independent predictor (new, active and vigorous) and a random effect for the 206 individuals in the sample. The inclusion of the random effect for individual enables us to provide statistical validation of the significance of the social and linguistic factors in the model over and above the varying behaviours of these individuals (Tagliamonte & Baayen, 2012:146). This model shows the relative influence of the linguistic factors together with the influence of community/age cohort. We include as predictors each of the internal factors as well as their interactions with community/age. If the interaction of community/age cohort and any of the internal predictors is significant, then we will take this as evidence that the predictor is changing over time.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: dep.var ~ COMMUNITY.AGE.1 + PROXIMITY + ANIMACY.PERSON + SENTENCE.TYPE +
CLAUSE.TYPE + COMMUNITY.AGE.1 * ANIMACY.PERSON + COMMUNITY.AGE.1 *      SENTENCE.TYPE +
COMMUNITY.AGE.1 * CLAUSE.TYPE + COMMUNITY.AGE.1 *      SENTENCE.TYPE + COMMUNITY.AGE.1 *
PROXIMITY + (1 | name)
   Data: fut
  AIC  BIC logLik deviance
 4640 4802  -2295     4590
Random effects:
 Groups Name        Variance Std.Dev.
 name   (Intercept) 0.29743  0.54538
Number of obs: 4764, groups: name, 206

Fixed effects:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                               -3.10071    0.40429  -7.669 1.73e-14 ***

COMMUNITY.AGE.1incipient
COMMUNITY.AGE.1new                         0.76851    0.43330   1.774 0.076127 .
COMMUNITY.AGE.1vigorous                    2.44303    0.50793   4.810 1.51e-06 ***

PROXIMITYfar
PROXIMITYnear                              0.87637    0.39307   2.230 0.025777 *
PROXIMITYno reference                      0.76279    0.37404   2.039 0.041416 *

ANIMACY.PERSON1stperson
ANIMACY.PERSONanimate                     -0.07807    0.21189  -0.368 0.712544
ANIMACY.PERSONinanimate                   -0.22781    0.33500  -0.680 0.496480

SENTENCE.TYPEdeclarative
SENTENCE.TYPEinterrogative                 1.45820    0.34722   4.200 2.67e-05 ***
SENTENCE.TYPEnegative                      1.28989    0.24931   5.174 2.29e-07 ***

CLAUSE.TYPEmain
CLAUSE.TYPEsubordinate                     0.80923    0.23628   3.425 0.000615 ***
```

```
COMMUNITY.AGE.1incipient:ANIMACY.PERSONanimate
COMMUNITY.AGE.1new:ANIMACY.PERSONanimate             0.42597    0.23690   1.798 0.072163 .
COMMUNITY.AGE.1vigorous:ANIMACY.PERSONanimate        0.71292    0.32890   2.168 0.030192 *

COMMUNITY.AGE.1incipient:ANIMACY.PERSONinanimate
COMMUNITY.AGE.1new:ANIMACY.PERSONinanimate           0.64006    0.36408   1.758 0.078742 .
COMMUNITY.AGE.1vigorous:ANIMACY.PERSONinanimate      0.91472    0.52112   1.755 0.079211 .

COMMUNITY.AGE.1incipient:SENTENCE.TYPEinterrogative
COMMUNITY.AGE.1new:SENTENCE.TYPEinterrogative       -0.28161    0.40224  -0.700 0.483854
COMMUNITY.AGE.1vigorous:SENTENCE.TYPEinterrogative -0.06481    0.86416  -0.075 0.940216

COMMUNITY.AGE.1incipient:SENTENCE.TYPEnegative
COMMUNITY.AGE.1new:SENTENCE.TYPEnegative            -0.96252    0.28451  -3.383 0.000717 ***
COMMUNITY.AGE.1vigorous:SENTENCE.TYPEnegative       -1.89334    0.53401  -3.545 0.000392 ***

COMMUNITY.AGE.1incipient:CLAUSE.TYPEsubordinate
COMMUNITY.AGE.1new:CLAUSE.TYPEsubordinate           -0.13701    0.25669  -0.534 0.593521
COMMUNITY.AGE.1vigorous:CLAUSE.TYPEsubordinate       0.57171    0.47912   1.193 0.232769

COMMUNITY.AGE.1incipient
COMMUNITY.AGE.1new:PROXIMITYnear                    -0.22258    0.41902  -0.531 0.595278
COMMUNITY.AGE.1vigorous:PROXIMITYnear               -1.08141    0.48529  -2.228 0.025854 *

COMMUNITY.AGE.1incipient:PROXIMITYno reference
COMMUNITY.AGE.1new:PROXIMITYno reference            -0.25391    0.39845  -0.637 0.523967
COMMUNITY.AGE.1vigorous:PROXIMITYno reference       -0.97709    0.48882  -1.999 0.045623 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 6: Generalized linear mixed model – all communities/age cohorts

Following Baayen (2008) we will use the index of concordance C to measure how well the model discriminates between the *be going to* and *will* variants. The Index of Concordance value of this model is .74. When C = .5 classification performance is at chance while values higher than .8 are considered very good. The p values are progressive for the community/age predictor in the order: incipient ($p > .01$), active (new) ($p > 0.001$), vigorous ($p > 0.001$), revealing that this partitioning of the data is highly significant and each one distinct from the other. Looking at the fixed predictors first, three of the four known influences on the use of *be going to* are statistically significant overall. The most important of these is the type of sentence, followed by type of clause and proximity. Animacy/person is not significant. However, the tests for interaction of these predictors by community/age all reach significance, exposing a significant reorganization of linguistic factors across communities, which we will interpret as an indication of change. Indeed, the extent of interaction here may explain the modest C value for this model. Among the interactions, the most dramatic is that between COMMUNITY/AGE and SENTENCE TYPE. What mechanism underlies this?

5.2 Animacy/grammatical person and sentence type

Figures 5, 6 and 7 cross-tabulate the marginal data for animacy/grammatical person and sentence type in order to clearly display the patterns in incipient, active and vigorous *be going to* situations (see Appendix A for the marginal data for the total data set).

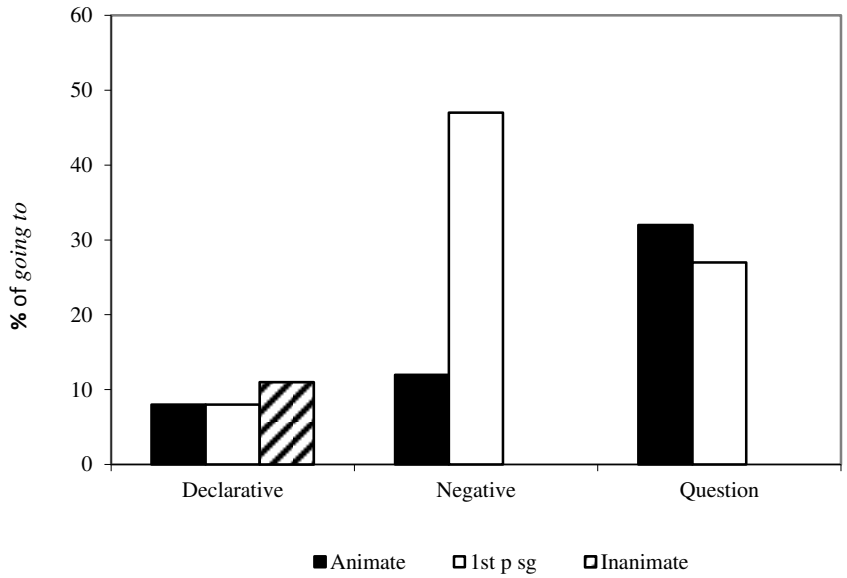

Figure 5: *Incipient* be going to

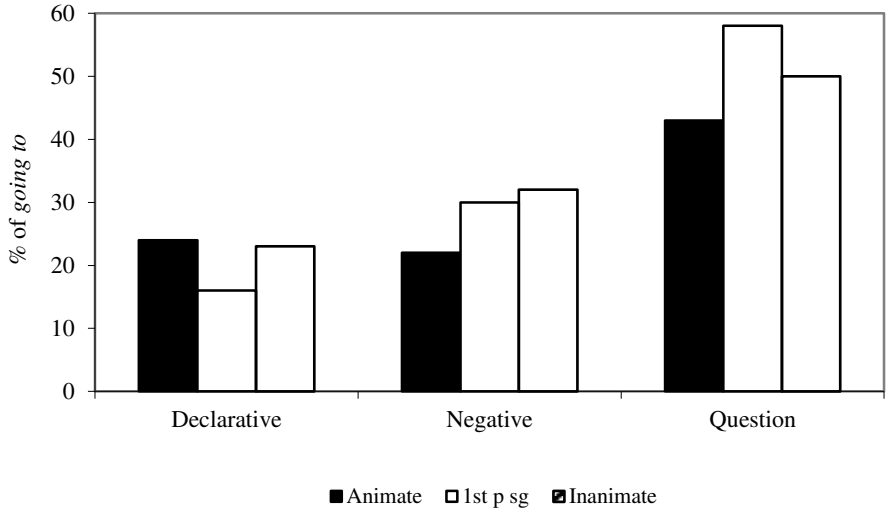

Figure 6: *Active* be going to

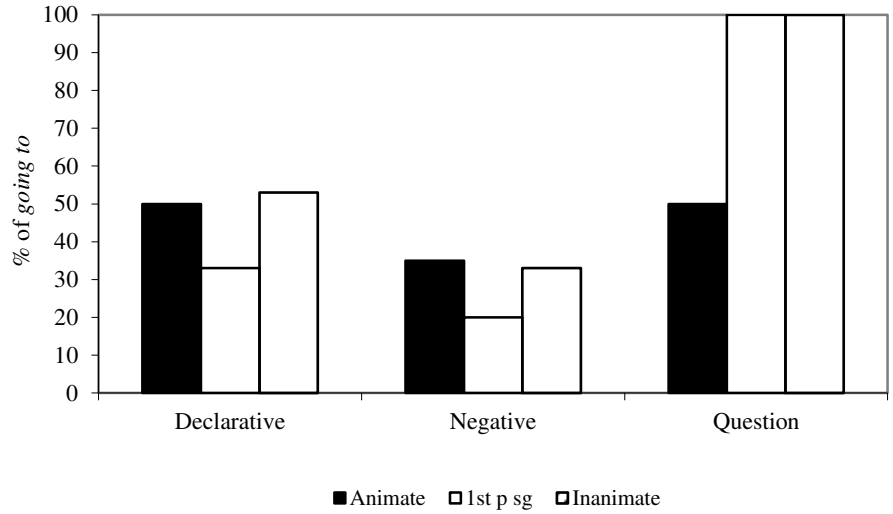

Figure 7: *Vigorous* be going to

In incipient *be going to* situations (Figure 5), 1st person negatives and questions have the highest frequency of *be going to* (47%), as in (19), (27%), as in (20) .

(19) I*'m nae gan* to argue about that! (BCK/7)

(20) What *am* I *gonna* do? (WHT/m)

Contrast this with the active *be going to* situation in Figure 6, where 1st person negatives do not stand apart questions are highly disposed to the incoming form. In Figure 7, the vigorous situation, *be going to* is as likely in declaratives as negatives and questions. Note that although 1st p. sg. and inanimates appear to have accelerated rates, these bars represent only 1 token each and so cannot be definitive. The main point here is that declaratives — the most ubiquitous context — is now hospitable to the incoming form. These divergent patterns from one community type to another explain the strong interactions in the amalgamated model.

5.3 Incipient vs. active vs. vigorous

With the shifting patterns of *be going to* in mind, we will now focus on the system in each community/age cohort where it is apparent that systemic reorganization of the internal predictors is in progress. To obviate problems arising from small cell sizes we restrict the models to the fixed predictors, the critical interaction between sentence type and animacy/grammatical person and a random effect of individual. Tables 7, 8 and 9 present the results.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: dep.var ~ ANIMACY.PERSON + SENTENCE.TYPE + CLAUSE.TYPE + PROXIMITY +
ANIMACY.PERSON * SENTENCE.TYPE + (1 | name)
   Data: incipient
   AIC   BIC logLik deviance
 810.5 875.7 -392.3    784.5
Random effects:
 Groups Name        Variance Std.Dev.
 name   (Intercept) 0.37002  0.60829
Number of obs: 1115, groups: name, 50

Fixed effects:
                                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                                      -3.3522     0.4272  -7.846 4.29e-15 ***

ANIMCACYPERSON1stperson
ANIMACY.PERSONanimate                             0.4078     0.2672   1.526  0.12700
ANIMACY.PERSONinanimate                           0.2575     0.3631   0.709  0.47823

SENTENCE.TYPEdeclarative
SENTENCE.TYPEinterrogative                        1.5554     0.7248   2.146  0.03187 *
SENTENCE.TYPEnegative                             2.5445     0.3716   6.848 7.51e-12 ***

CLAUSE.TYPEmain
CLAUSE.TYPEsubordinate                            0.7376     0.2415   3.054  0.00226 **

PROXIMITYfar
PROXIMITYnear                                     0.7633     0.3990   1.913  0.05572 .
PROXIMITYno reference                             0.7174     0.3818   1.879  0.06021 .

ANIMACY.PERSONanimate:SENTENCE.TYPEinterrogative    -0.1294     0.8299  -0.156  0.87604
ANIMACY.PERSONinanimate:SENTENCE.TYPEinterrogative -14.8350  2049.3654  -0.007  0.99422
ANIMACY.PERSONanimate:SENTENCE.TYPEnegative          -2.2163     0.5417  -4.091 4.29e-05 ***
ANIMACY.PERSONinanimate:SENTENCE.TYPEnegative      -15.4608   784.3936  -0.020  0.98427
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 7: Incipient *be going to*

Table 7 shows the incipient *be going to* situations. The Index of Concordance value of this model is .75. The strongest predictor is sentence type. Both interrogatives and negatives favour *be going to*, particularly negatives ($p > .001$) while declaratives are conservative, favouring *will*. Clause type is also significant with subordinate clauses favouring *be going to* ($p > .01$), whereas main clauses favour *will*. Near future contexts and those with no temporal reference exhibit a near equal and modest favouring effect for *be going to* ($p > .1$) while those that are far future disfavour *be going to*. Animacy/grammatical person is not significant as a

fixed effect in the model; however notice that the interaction between animacy/grammatical person and sentence type is dramatic: animate negatives that are 1st person strongly favour *be going to* (see also Figure 5).

```
Generalized linear mixed model fit by the Laplace approximation
Formula: dep.var ~ ANIMACY.PERSON + SENTENCE.TYPE + CLAUSE.TYPE + PROXIMITY +
ANIMACY.PERSON * SENTENCE.TYPE + (1 | name)
   Data: new
  AIC  BIC logLik deviance
 3339 3418  -1656     3313
Random effects:
 Groups Name        Variance Std.Dev.
 name   (Intercept) 0.29552  0.54362
Number of obs: 3291, groups: name, 141

Fixed effects:
                                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                                      -2.4412     0.1608 -15.179  < 2e-16 ***

ANIMACY.PERSON1stperson
ANIMACY.PERSONanimate                             0.5108     0.1163   4.391 1.13e-05 ***
ANIMACY.PERSONinanimate                           0.4866     0.1549   3.141 0.001681 **

SENTENCE.TYPEdeclarative
SENTENCE.TYPEinterrogative                        2.2011     0.6211   3.544 0.000394 ***
SENTENCE.TYPEnegative                             0.9218     0.2376   3.880 0.000105 ***

CLAUSE.TYPEmain
CLAUSE.TYPEsubordinate                            0.6712     0.1005   6.679 2.40e-11 ***

PROXIMITYfar
PROXIMITYnear                                     0.6767     0.1457   4.646 3.39e-06 ***
PROXIMITYno reference                             0.5092     0.1377   3.697 0.000218 ***

ANIMACY.PERSONanimate:SENTENCE.TYPEinterrogative    -1.2539     0.6620  -1.894 0.058199 .
ANIMACY.PERSONinanimate:SENTENCE.TYPEinterrogative  -0.7347     0.8480  -0.866 0.386317
ANIMACY.PERSONanimate:SENTENCE.TYPEnegative         -0.9570     0.3014  -3.175 0.001497 **
ANIMACY.PERSONinanimate:SENTENCE.TYPEnegative       -0.4083     0.4401  -0.928 0.353528
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 8: Active *be going to*

Table 8 shows the active *be going to* situations. The Index of Concordance value of this model is .71. The strongest predictor in this situation is type of sentence. The reference level is dramatically set apart from interrogatives as well as negatives, although notice the weakening of this effect compared to the vigorous situation in Table 7. Proximity is another very strong predictor. In this case, the reference level ‘far’ is significantly different from both near and no reference contexts. Similarly, animacy/grammatical person is significant, with first person subjects significantly differentiated from all other subject types. Moreover, the interaction factor group shows that declaratives remain distinct from interrogatives and negatives.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: dep.var ~ ANIMACY.PERSON + SENTENCE.TYPE + CLAUSE.TYPE + PROXIMITY +
ANIMACY.PERSON * SENTENCE.TYPE + (1 | name)
   Data: vigorous
   AIC   BIC logLik deviance
 478.5 528.9 -226.2    452.5
Random effects:
 Groups Name        Variance Std.Dev.
 name   (Intercept) 0.26249  0.51234
Number of obs: 358, groups: name, 15

Fixed effects:
                                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                                       0.68186    0.30833   2.211  0.02700 *

ANIMACY.PERSON1st person
ANIMACY.PERSONanimate                            -0.67089    0.25865  -2.594  0.00949 **
ANIMACY.PERSONinanimate                          -0.64351    0.41242  -1.560  0.11868

SENTENCE.TYPEdeclarative
SENTENCE.TYPEinterrogative                      -15.48227 1264.19457  -0.012  0.99023
SENTENCE.TYPEnegative                            0.65115    1.17900   0.552  0.58075

CLAUSE.TYPEmain
CLAUSE.TYPEsubordinate                           -1.36865    0.41675  -3.284  0.00102 **

PROXIMITYfar
PROXIMITYnear                                     0.19229    0.28606   0.672  0.50145
PROXIMITYno reference                             0.19269    0.31482   0.612  0.54050

ANIMACY.PERSONanimate:SENTENCE.TYPEinterrogative 14.83614 1264.19486   0.012  0.99064
ANIMACY.PERSONinanimate:SENTENCE.TYPEinterrogative  0.53690 1582.47390   0.000  0.99973
ANIMACY.PERSONanimate:SENTENCE.TYPEnegative      -0.06710    1.30846  -0.051  0.95910
ANIMACY.PERSONinanimate:SENTENCE.TYPEnegative     0.03086    1.76463   0.017  0.98605
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 9: Vigorous *be going to*

Table 9 shows the vigorous *be going to* situation. The Index of Concordance value of this model is .71. There is a shift in the predictors that are significant and to what degree. First, sentence type is no longer significant. Secondly, clause type has decreased in strength. Third, the effect of proximity levels is different. While this was one of the strongest effects in the incipient situation, yet here there is no longer a statistically significant effect. These results jibe with the idea that as *be going to* grammaticalizes it becomes more “colorless” (Joos, 1964:23) or neutral (Brisard, 2001). Finally, animacy/grammatical person shows a clear contrast between 1st person and animate subjects types, as in the active *be going to* situations; however, now it has expanded into inanimate contexts since there is no statistically significant contrast between first person and inanimates.

In sum, we can infer three distinct changes in the use of *be going to* across incipient, active and vigorous stages from these data. First, when *be going to* starts to be used for future

temporal reference, it does so with first person questions and negatives. Negatives are most hospitable, followed by interrogatives. Declarative sentences are only gradually infiltrated by *be going to*. As it increases in frequency, this strong tendency changes. The conservative nature of 1st person, however, endures across all community/age groups. Second, the effect of proximity changes dramatically: the association of *be going to* with no reference and near reference futures levels out as *be going to* expands into far future contexts. Further, it appears that in incipient contexts the proximity effect is not necessarily "near" as previous research has suggested, but simply "not far". Both "near" and "no reference" contexts favour incoming *be going to.* Far future contexts stand apart; in these contexts *be going to* is rare. Finally, the effect of clause type is stable across all contexts. Thus, in contrast to previous studies, in our data there is no development towards increasing use of *be going to* in main clauses.

We now review each of the predictors in turn according to statistical significance, constraint ranking and relative weights of the factors. The complexity of this data structure and the cross-variety nuances of the sample make it particularly important to base our interpretations on a convergence of all the evidence available in the analyses (see Tagliamonte, 2007:204).

The effect of sentence type shows the most dramatic change across the community/age cohorts. First, the constraint re-organizes: among the older speakers both negatives and interrogatives favour *be going to* but the favouring effect of negatives declines from highly favouring to irrelevant among the York younger speakers. Second, the strength of the constraint decreases from oldest to youngest speakers. This adds to the earlier comparison (Tables 3-4) in elucidating the change in this constraint. Taken together with Szmrecsanyi's (2003) observation[29] that British varieties favour *I won't* (or *I'll not)* over *I'm not going to* leads us to surmise that his data represent a more conservative variety and/or stage in the development of *be going to*.

[29] Szmrecsanyi's (2003) data came from the British National Corpus.

The effect of clause type is significant in the amalgamated data with no interaction by community/age cohort and is significant in each of the individual community/age cohort models. The fact that *be going to* is said to have arisen first in subordinate clauses suggests that this is a longtime retention of earlier patterns.

The effect of proximity is confirmed but it involves an effect of far future rather than near future as has often been suggested in the literature. The data make it evident that far future contexts have a tendency to resist *be going to*. Among the Buckie older speakers, for example, they are never marked by *be going to*; however far futures takes on more *be going to* from incipient to active situations. In the vigorous situation there is no difference across temporal reference contexts and the constraint is not significant.

The effect of animacy/grammatical person incrementally changes across cohorts (as also visible in Figures 3-7). In Figure 3, the frequency of *be going to* rises across generations in apparent time. In Figure 4 the conditional inference tree splits the populations by age in virtually every splitting of the data. In Figures 5-7, which probes the animacy/grammatical person predictors more deeply, we see changes from incipient to active to vigorous. In the youngest group in York 1st person singular still stands apart; however, only animates remain statistically different, which establishes the developmental extension of *be going to* into inanimates. The original association of *be going to* with animate subjects, particularly 1st person, has shifted to a strong effect for the use of *'ll* for first person.[30]

Taken together, we suggest that the results for the oldest generation across communities and the generational trends within and across communities offer new insight into the early grammatical development of *be going to*. We propose the following pathway of grammaticalization. First person singular negatives and questions play a key role in the early stage. In contrast to earlier accounts which put special emphasis on 1st person animates capable of movement, we suggest that personal intention, which is particularly strong in negatives and questions, played the more important role. This is supported by: 1) examples of

[30] The York young speakers clearly show that this is an effect of *'ll* rather than 'will' more generally. The contracted variant represents 60.6% of the 1st person contexts compared to *be going to* 31.6% and *will* 7.7%.

be going to in the Helsinki Corpus which have intention as a highly foregrounded meaning; and 2) early examples of the verb *intend* from the same time period which collocate with 1st person singular (see example (3)c). These few examples from the historical record show that the intention reading was present from the earliest uses. The literature identifies subordinate clauses as the early entry point for *be going to.* However, our findings suggest that contexts of 1st person intention and interrogatives may have been early favouring contexts for *be going to* as well, perhaps more so. It is not solely the results in Table 7 that suggest this but also the fact that these varieties represent localities where *be going to* is apparently just beginning to encroach on the future temporal reference system. Studies with higher rates of *be going to* (such as in North America) embody the system at a point in its trajectory of development where this early penetration into the system is long past. The resistance of *be going to* to 1st person singular subjects more generally, as visible in the contrast between 1st person and animates in the vigorous context ($p < .01$), must be a later development linked to prescriptive norms or to the increased strength of the collocation *I'll.* This is consistent with the characteristics of older grammatical morphemes, which tend to be fused and short (Bybee et al., 1994:47). Finally, extension of *be going to* to inanimates and far future readings is a noticeably progressive development, visible in the shifting values of coefficients and relative strength of this factor in the predictor (animacy/grammatical person). From the active to the vigorous situation, inanimate contexts have more *be going to*, to the point where they are not distinguished from 1st person contexts. Similarly, far future readings go from being distinguished from the other future readings to the York young data, where there is no statistically significant difference among types.

If we step back and take an over-arching look at the results, we see that there are relatively dramatic differences between the old Buckie speakers and the young York speakers — in terms of frequency as well as predictors' significance, constraint ranking and strength of effect. Following this through, we can now make the observation that where we observe dramatic changes in frequency (13% vs. 22% vs. 41%) re-organization of constraints is paramount. Thus, grammaticalization certainly involves increasing frequency as well as shifting weights of constraints. However, when frequency differences are greatest, they

correlate with a brisk reorganization in certain areas of the grammar. In this way, the system advances with incremental, often minute changes (as visible in the shifting strength of factors), and appears to transition from one (stable) pattern to another in the context of substantial acceleration of the incoming form. These developments are not necessarily expected. Many changes advance by moving forward at the *same* rate in all contexts, i.e. the Constant Rate Effect (Kroch 1989, 2003), for example, rise of *do* support Ellegård 1953). In grammaticalization, however, constraints (patterns) necessarily re-organize via analogical extension, leveling, pragmatic strengthening and the like. We suggest that the two patterns should, in theory, be visible in empirical data, permitting analysts to infer what type of change is taking place (see Tagliamonte 2012:83-91). In the case at hand, the interpretation points to grammatical change.

In sum, by examining varieties at varying stages in the process of change we have been able to uncover what we believe are important insights into the evolutionary pathway of *be going to* as it rises in frequency and develops new patterns. The two-tiered approach of examining different communities and different generations has allowed us to establish subtle shifts that would not have been apparent otherwise and the elusive sparks of acceleration that push the system forward.

6. CONCLUSIONS

Regional dialects (space) and speakers of varying ages (apparent time) have offered a vernacular, community-based perspective into the history of the grammaticalization of *be going to* in English across the entire first half of the S curve.

Our findings suggest that some of the well-known and often-attested constraints on *be going to* must have been present at the earliest stages of grammaticalization since we have found them to be significant in situations where *be going to* has just begun to make headway into the grammar. These constraints include: 1) the strong effect of *be going to* with questions; 2) the preponderance of *be going to* in subordinate clauses; and 3) the inhospitality of *be going to* to far future meaning. This provides yet another demonstration that

grammaticalization does not require a certain threshold in frequency for the operation of relevant constraints. Moreover, it confirms the idea that "tenacious patterns of distribution" can last across centuries and shape grammatical variation over the long term (Torres-Cacoullos & Walker, 2009:323). The composite of meanings noted in the literature on the *go* future involving movement and/or intention with animate subjects in embedded clauses is a case in point. This is a virtually un-extricable set of meanings that can be found in the earliest uses of *be going to*, whether they are clearly future or not, right up to the present-day. Moreover, we have documented the well known patterns of the grammaticalization of *be going to* across dialects that have never been studied for this feature before. In this way our findings bolster the idea of universal pathways of grammaticalization where meanings are shaped and smoothed and sculpted out of lexical and discourse patterns (Sankoff & Brown, 1976, Poplack & Tagliamonte, 1999, Bybee, 2006). Some of the associations we have uncovered precisely embody these kinds of systemic processes. Developments such as extension and levelling take place over the long term and in incremental steps along the way. For example the use of *be going to* starts out in interrogative and negative sentences, but then extends to declaratives. Similarly, the use of *be going to* for near and no reference is extending to far future contexts. In both these cases, *be going to* spreads out from an association with a particular context to a more general set. The same trends are apparent in the development of other tense/aspect systems such as the extension of hodiernal past into the general past temporal reference system (Dahl, 1984) or an anterior perfect generalizing to a perfective (Schwenter, 1994). The development of the contracted form of *will* as *'ll* which is fused with 1st person singular is another over-arching trend. As far as our data are concerned, this is a relatively recent development and, at least in part, must be a more general process in the aging of grammatical morphemes. Taken together, all these widely diverging patterns of change — some reactionary, some revolutionary — reveal the complex nature of grammaticalization generally. These complimentary impacts on the system may be part of the explanation for the longitudinal layering of *be going to* and *will* in contemporary English.

REFERENCES

Santa Barbara Corpus of Spoken American English. http://www.linguistics.ucsb.edu/research/santa-barbara-corpus. Accessed 11-8-13 11:06am.

The Helsinki Corpus of English Texts. Helsinki: Department of English, University of Helsinki. http://icame.uib.no/hc/. Accessed 11-8-13 11:06am.

Aceto, Michael. 1998. A new creole future tense marker emerges in the Panamanian West Indies. *American Speech* 73(1), 29-43.

Anderson, Eric W. 1979. The development of the Romance future tense: morphologization and a tendency toward analycity. *Papers in Romance* 1, 21-35.

Arnovick, Leslie K. 1990. *The development of future constructions in English*. New York: Peter Lang.

Baayen, Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Berglund, Ylva. 1997. Future in present-day English: Corpus-based evidence on the rivalry of expressions. *ICAME Journal* 21, 7-19.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.

Brinton, Laurel & Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.

Brisard, Frank. 2001. *Be going to*: An exercise in grounding. *Journal of Linguistics* 37, 251-85.

Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In Joseph, B. D. & R. D. Janda (eds.), *The handbook of historical linguistics*, 602-23. Oxford: Blackwell.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(2), 711-33.

Bybee, Joan L. & William Pagliuca. 1987. The evolution of future meaning. In Ramat, A. G., O. Carruba & G. Bernini (eds.), *Papers from the 7th international conference on historical linguistics*, 107-22. Amsterdam and Philadelphia: John Benjamins.

Bybee, Joan L., William Pagliuca & Revere D. Perkins. 1991. Back to the future. In Traugott, E. C. & B. Heine (eds.), *Approaches to grammaticalization. Volume II*, 19-58. Amsterdam and Philadelphia: John Benjamins.

Bybee, Joan L., Revere D. Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Close, R. A. (1977). Some observations on the meaning and function of verb phrases having future reference. In Bald, W.-D. & Ilson, R. (Eds.), *Studies in English usage: the resources of a present-day English corpus of linguistic analysis*. Frankfurt: Peter Lang. 125-156.

Crystal, David. 1986. I shall and I will. *English Today* 5(Jan-Mar), 42-4.
Curme, George O. 1913. Has English a future tense? *Journal of English and Germanic Philology* 12(4), 515-39.
Dahl, Östen (1984). Temporal distance: Remoteness distinctions in tense-aspect systems. In Butterworth, B., Comrie, B. & Dahl, O. (Eds.), *Explanations for Language Universals.* Berlin: Mouton de Gruyter. 105-122.
Danchev, Andrei & Merja Kytö. 1994. The construction *be going to* + infinitive in Early Modern English. In Kastovsky, D. (ed.), *Studies in Early Modern English*, 59-77. Berlin: Mouton de Gruyter.
Danchev, Andrei & Merja Kytö. 2002. The *go*-futures in English and French viewed as an areal feature. *Nowele* 40, 29–60.
Disney, Steve (2009). The grammaticalisation of 'BE going to'. *Newcastle working papers in Linguistics* 15: 63-81.
Elson, M. 1988. The morphology of the future and conditional in Spanish. *Revue Roumaine de linguistique* 33, 3-12.
Ellegård, Alvar (1953). *The auxiliary do: the establishment and regulation of its use in English.* Stockholm: Almqvist and Wikwell.
Fleischman, Suzanne. 1982. *The future in thought and language: Diachronic evidence from Romance*. Cambridge: Cambridge University Press.
Fries, Charles Carpenter (1925). *The periphrastic future with "shall" and "will" in modern English.* Baltimore: The Modern Linguistic Association of America.
Fries, Charles Carpenter. 1927. The expression of the future. *Language* 3(2), 87-95.
Fries, Charles Carpenter. 1940. *American English grammar*. New York: Appleton, Century, Crofts.
Gagnon, Sylvie. 1990. Future proche, futur éloigné: Problème terminologique et didactique. *Langues et Linguistique* 16, 253-8.
Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.
Grancrov, Hristo. 1992. Some peculiarities in the use of future periphrases with *shall*/*will* in the early Modern English period 15th-17th century. *Contrastive Linguistics* 17(3), 99-104.
Haegeman, Lilliane. 1981. *The use of* will *and the expression of futurity in Present-day British English.* Part 1*: Future time expression in a corpus of Standard British English*. D.Phil. Dissertation, University College, London.
Hall, R. M. R. & Beatrice L. Hall. 1970. A Note on 'will' vs. 'going to' *Linguistic Inquiry* 1, 138-9.
Harada, Shigeo (1958). The "*be+going+to*"-infinitive form in Shakespeare. In Kazuo, A. (Ed.), *Studies in English Grammar and Linguistics. A Miscellany in Honour of Takanobu Otsuka.* Tokyo: Kenkyusha. 317-322.
Harner, Lorraine. 1976. Children's understanding of linguistic reference to past and future. *Journal of psycholinguistic research* 5(1), 65-84.
Harner, Lorraine. 1980. Immediacy and certainty: Factors in understanding future reference. *Journal of Child Language* 9(2), 115-24.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Hilpert, Martin. 2008. *Germanic Future Constructions: A usage-based approach to language change*. Amsterdam and Philadelphia: John Benjamins.

Hoffmann, Sebastian (2005). *Grammaticalization and English Complex Prepositions. A Corpus-Based Study.* London: Routledge.

Hopper, Paul J. 1991. On some principles of grammaticization. In Traugott, E. C. & B. Heine (eds.), *Approaches to grammaticalization,* Volume 1: *Focus on theoretical and methodological issues*, 17-35. Amsterdam and Philadelphia: John Benjamins.

Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*, (Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

Howe, Neil & William Strauss. 1991. *Generations: The history of America's future 1584-2069*. New York: William Morrow and Co. Inc.

Joos, M. 1964. *The English verb: Form and meaning*. Madison and Milwaukee: University of Wisconsin Press.

Kroch, Anthony S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3): 199-244.

Kroch, Anthony (2003). Syntactic change. In Baltin, M. & Collins, C. (Eds.), *The Handbook of Contemporary Syntactic Theory.* Malden and Oxford: Blackwell.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Leech, Geoffrey, Hundt, Marianne, Mair, Christian, & Smith, Nicholas. (2009). *Change in contemporary English: A grammatikcal study.* Cambridge: Cambridge University Press.

Luebke, W. F. 1929. The analytic future in contemporary American fiction. *Modern Philology* 26, 451-7.

Mair, Christian. 1997. The spread of the *going-to*-future in written English: A corpus-based investigation into language change in progress. In Hickey, R. & S. Puppel (eds.), *Language history and linguistic modelling*, *Volume II*, 1537-43. Berlin: Mouton de Gruyter.

Mair, Christian. 2004. Corpus linguistics and grammaticalisation theory: Statistics, frequencies and beyond. In Lindqvist, H. & C. Mair (eds.), *Corpus Approaches to Grammaticalization in English*, 121-50. Amsterdam: Benjamins.

Mair, Christian (2006). *Twentieth century English: History, variation and standardization.* Cambridge: Cambridge University Press.

Mossé, Fernand. 1952. *A handbook of Middle English*. Baltimore: John Hopkins University Press.

Myhill, John (1988). A quantitative study of future tense marking in Spanish. In Ferrara, K., Brown, B., Walters, K. & Baugh, J. (Eds.), *Linguistic change and contact.* Austin: University of Texas. 263-272.

Nehls, Dietrich. 1988. Modality and the expression of future time in English. *International Review of Applied Linguistics* 26(4), 295-307.

Nevalainen, Terttu & Raumolin-Brunberg, Helena (1996). *Sociolinguistics and*

language history: Studies based on the corpus of Early English correspondence. Amsterdam and Atlanta: Rodopi.
Nieuwint, Pieter. 1986. Present and future in conditional protases. *Linguistics* 24(2), 371-92.
Ofuani, Ogo A. 1981. Future time expression in Nigerian Pidgin. *International journal of human communication* 14(3), 309-26.
Palmer, Frank R. 1974. *The English verb*, 2 edn. (Robins, R. H.
Leech, Geoffrey N. (1987). *Meaning and the English verb.* New York: Longman.
Pérez, Aveline. 1990. Time in Motion: Grammaticalisation of the *be going to* construction in English. *La Trobe University Working Papers in Linguistics* 3, 49-64.
Poplack, Shana & Elisabete Malvar. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19, 121–69.
Poplack, Shana & Sali A. Tagliamonte. 1999. The grammaticalization of *going to* in (African American) English. *Language Variation and Change* 11(3), 315-42.
Poplack, Shana & Danielle Turpin. 1999. Does the FUTUR have a future in (Canadian) French? *Probus* 11, 133-64.
Quirk, Randolph, Sidney Greenbaum, Geoffry Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman.
Roberts, Ian. 1992-3. A formal account of grammaticalisation in the history of Romance futures. *Folia Linguistica Historica* 13/1-2, 219-58.
Robertson, T. A. & John Graham. 1952/1991. *Grammar and Usage of the Shetland Dialect*. Lerwick: Shetland Times Ltd.
Roy, Joseph. 2007. *Peering into the future: The emergence of* going to *in the 18th century*. Presented at the New Ways of Analyzing Variation 36, University of Pennsylvania, Philadelphia.
Royster, Jane F. & John M. Steadman. 1923/1968. The "going-to" future. In, *Manly Anniversary Studies in Languages and Literature*, 394-403. Freeport, N.Y.: Books for Libraries Press.
Sankoff, Gillian & Penelope Brown. 1976. The origins of syntax in discourse. *Language* 52(3), 631-66.
Schwegler, Armin. 1992. Future and conditional in Palenquero. *Journal of Pidgin and Creole Languages* 7(2), 223-59.
Schwenter, Scott. 1994. The grammaticalization of an anterior in progress: Evidence from a peninsular Spanish dialect. *Studies in Language* 18(1), 71-111.
Smith, Jennifer & Mercedes Durham. 2011. A Tipping Point in Dialect Obsolescence? Change across the Generations in Lerwick, Shetland. *Journal of Sociolinguistics* 15(2), 197-225.
Smith, Jennifer & Mercedes Durham. 2012. Bidialectalism or dialect death? Explaining generational change in the Shetland Islands, Scotland. *American Speech* 87(1), 57-88.
Strobl, C., J. Malley & G. Tutz. 2009. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classifi-

cation and Regression Trees, Bagging, and Random Forests. *Psychological methods,* 14(4), 323-48.
Szmrecsanyi, Benedikt. 2003. 'Be going to' versus 'will/shall'. Does syntax matter? *Journal of English Linguistics* 31, 295-323.
Tagliamonte, Sali A. 2002. Comparative sociolinguistics. In Chambers, J. K., P. Trudgill & N. Schilling-Estes (eds.), *Handbook of language variation and change*, 729-63. Malden and Oxford: Blackwell Publishers.
Tagliamonte, Sali A. 2007. Quantitative analysis. In Bayley, R. & L. Ceil (eds.), *Sociolinguistic variation: Theory, methods, and applications, Dedicated to Walt Wolfram*. Cambridge: Cambridge University Press.
Tagliamonte, Sali A. 2011. Variation as a window on universals. In Siemund, P. (ed.), *Linguistic Universals and Language Variation* Berlin/New York: Mouton de Gruyter.
Tagliamonte, Sali A. (2012). *Variationist Sociolinguistics: Change, observation, interpretation.* Malden and Oxford: Wiley-Blackwell.
Tagliamonte, Sali A. 2013a. The verb phrase in contemporary Canadian English. In Aarts, B., J. Close, G. Leech & S. A. Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*, 133-54. Cambridge: Cambridge University Press.
Tagliamonte, Sali A. 2013b. *Roots of English: Exploring the history of dialects.* Cambridge: Cambridge University Press.
Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2).
Tagliamonte, Sali A. & Alexandra D'Arcy. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85(1), 58-108.
Tagliamonte, Sali A. & Jennifer Smith. 2006. Layering, change and a twist of fate: Deontic modality in dialects of English. *Diachronica* 23(2), 341-80.
Tagliamonte, Sali A., Jennifer Smith & Helen Lawrence. 2005. English dialects in the British Isles in cross-variety perspective: A base-line for future research. In Filppula, M., J. Klemola, M. Palander & E. Penttilä (eds.), *Dialects across borders: Selected papers from the 11th international conference on methods in dialectology, Joensuu, August 2002*, 87-117. Amsterdam and Philadelphia: John Benjamins.
Team, R Development Core. 2007. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing http://www.R-project.org.
Torres-Cacoullos, Rena & James A. Walker. 2009. The present of the English Future: Grammatical variation and collocations in discourse. *Language* 85(2), 321-54.
Trudgill, Peter J. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Languge in society* 3, 215-46.
Ultan, R. 1972. The nature of future tenses. *Working papers in language universals* 8, 55-100.

Wagner, Suzanne Evans & Gillian Sankoff. 2011. Age grading in the Montréal French inflected future. *Language Variation and Change* 23(3), 275-313.

Wales, M. L. 1983. The semantic distribution of "aller + infinitive" and the future tense in spoken French. *General Linguistics* 23(1), 19-28.

Wekker, H. C. 1976. *The expression of future time in contemporary British English*. Amsterdam: North Holland.

Williams, Christopher. 2013. Changes in the verb phrase in legislative language in English. In Aarts, B., J. Close, G. Leech & S. A. Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora,* 353-71. Cambridge: Cambridge University Press

Appendix A

	Declarative		Negative		Interrogative		Total N
	%	N	%	N	%	N	
Incipient *be going to*							
1st person	8	265	49	41	33	9	315
Animate	10	322	12	58	19	26	406
Inanimate	10	90	0	8	0	0	98
Vigorous *be going to*							
1st person	13	358	30	33	0	1	392
Animate	24	421	24	55	44	36	512
Inanimate	24	122	36	11	44	9	142

Cross-tabulation of animacy/grammatical person and sentence type in use of *be going to*