*English Language and Linguistics* 00.0: 1–41. © Cambridge University Press 2016
doi:10.1017/S136067431600023X

# Sociophonetic variation of *like* in British dialects: effects of function, context and predictability[1]

Q1

ERIK SCHLEEF
*University of Salzburg*

and

DANIELLE TURTON
*Newcastle University*

(Received 22 December 2014; revised 27 October 2015)

This study examines sociophonetic variation in different functions of *like* among adolescents in London and Edinburgh. It attempts to determine the factors that may explain this variation. Our results suggest that the function of *like* correlates primarily with contextual factors, rather than the phonetic factors of vowel quality, /l/ to vowel duration and /k/ realisation. In particular, the preceding and following segments and their bigram predictability emerge as highly significant, in addition to the boundary strength following *like*. In both London and Edinburgh, the vowel appears to be the only non-contextual feature that is sensitive to the function of *like*: quotative *be like* is more likely to be monophthongised than other functions of *like*. We argue that the more monophthongal nature of quotative *like* is due to the syntactic and prosodic context in which it occurs.

## 1 Introduction

The last three decades have witnessed the appearance of a large number of studies on the various discourse functions of *like* (e.g. Underhill 1988; Buchstaller 2004; D'Arcy 2006; Tagliamonte & D'Arcy 2007; Cheshire, Kerswill, Fox & Torgersen 2011; Durham *et al.* 2012). We have learned much about how *like*, particularly quotative *be like*, is constrained in several varieties of English and what happens to these constraints and social meanings of *like* once it enters a new variety (e.g. Buchstaller 2006; Buchstaller & D'Arcy 2009). While we assume generally that the lexical form of discourse *like* will be adapted into the linguistic system of any new variety it enters, so that in Southern English varieties in the UK it will appear most commonly as [laɪk], or for Scottish varieties [lʌik], we still know relatively little about the conversational phonetics of *like* in its various discourse functions in different varieties of English.

[1] This research was funded by the UK Arts and Humanities Research Council (AHRC, grant AH/K003674/1, Erik Schleef PI). We are grateful to Fernanda McDougall and Patrycja Strycharczuk for assisting ably with coding and corpus handling. We thank Josef Fruehwald for providing us with the Praat script for speech rate, and Constantine Lignos for his assistance in extracting bigram frequencies from the SUBTLEX corpus. We also thank Maciej Baranowski, Ricardo Bermúdez-Otero, Tine Breban, Jenny Cheshire, Yuni Kim, Laurel MacKenzie and the members of the University of Manchester Phonology reading group for their expert advice on the topic. Particular thanks go to two anonymous reviewers for providing us with extremely helpful, considerate and insightful comments, which have shaped this paper. We alone are responsible for any failings in this paper.

Drager's (2009, 2011) work on *like* in a New Zealand high school is a notable exception. While investigating three different functions of *like* (grammatical, quotative and discourse particle), she found that *like* does indeed have different phonetic realisations across functions. For example, quotative *like* tends to be less diphthongal than *like* used in other functions. Drager (2011) takes her finding of the systematically different realisation of *like* functions as evidence that these items are stored in the mind in such a way that functional distinctions are maintained, and that there must be a direct link between lemma-based and acoustically rich information.

In the UK too, there is anecdotal evidence of *like* reduction: in London English, *like* may be pronounced as monophthongal [la:k] or with final consonant reduction [laɪ], whereas in Edinburgh we find reduced forms such as [lɪk]. Thus, *like* is subject to different system-internal pressures in New Zealand, London and Edinburgh, which allows us to test Drager's argument further, whilst providing a regional comparison. The current study has two goals. It aims to (a) determine the nature of the sociophonetic variation of different functions of *like* among adolescents in London and Edinburgh, and (b) explore which factors may explain this variation. We will extend our investigative frame beyond the phonetic and contextual predictors investigated by Drager and also include factors, such as prosody and word probability, to uncover whether any reductive processes may be conditioned by the linguistic context in which *like* occurs rather than discourse function. Thus, the findings will enable us to reflect on the theoretical arguments proposed by Drager (2011). This is because further evidence of contextual conditioning of function-based reduction would undermine arguments in favour of function-based storage of phonetic detail in the mind. Based on the speech of teenagers in London and Edinburgh, we demonstrate that the phonetic makeup of *like* depends first and foremost on contextual factors.

## 2 *Like*: functions, forms, development and reduction

This section outlines the different functions of *like* and justifies the current study. More detail is provided on research that has investigated the phonetics of different functions of *like* as well as descriptions of /l/, /aɪ/ and /k/ realisations in London and Edinburgh. The section concludes with our hypotheses regarding the two questions raised above.

### *2.1* Like *and its functions*

Various forms of *like* can be differentiated. In order to make our data comparable with the majority of other studies conducted in the tradition of variationist sociolinguistics, we follow D'Arcy (2007: 392–5) in differentiating between these. The lexical item under investigation can, of course, be used in a variety of standard forms: as a verb (*I just love Ireland, I* ***like*** *the music and everything there*; Edinburgh 004, Jenna, 15), adverb (*our teacher looks* ***like*** *postman Pat*; Edinburgh 002, Debbie, 14), conjunction (*they pure talk to us* ***like*** *we're something scraped off their shoe*; Edinburgh 009, Skye, 14), noun (*the* ***likes***), and suffix, as in *massive-**like*** (London 023, Thomas, 13). With the exception of *like* used as a noun, all of these occur in our corpus, from which examples have been taken. We provide subcorpus (Edinburgh, London) and speaker information (speaker number, participant-selected pseudonym and age) following the example.

We are particularly interested in the vernacular forms and differentiate between approximative adverbs, discourse markers, discourse particles and quotative *be like,* as defined in D'Arcy (2007).

Approximative adverbs signal approximation and are substituted frequently with *about* or *around* (Schourup 1983: 30): *Um, I haven't been for* ***like*** *three years now* (London 013, April, 14).

Discourse markers mark discourse and information structure outside the clause: *I don't know, when you go through to like Glasgow and stuff.* ***Like*** *Glasgow's really nice and stuff but I'm kind of always glad I did like grow up in Edinburgh* (Edinburgh 015, Lucy, 14).

Discourse particles mark discourse and information structure within clauses. They are used for focus. They also tend to occur before new information that has just been introduced in the interaction (Dailey-O'Cain 2000): *Like we wouldn't really* ***like*** *go on a train to* ***like*** *Birmingham or something like that, it would just be* ***like*** *too far* (London 023, Thomas, 13).

Quotative *be like* conveys a sense of approximation and introduces reported speech (*So she* ***was like****, aye, it looks like cellulitis*; Edinburgh 011, Terrance Charles Desmond, 16), thought (*But they try and speak Scottish it's stupid, it'****s like*** *go and talk your ain language*; Edinburgh 009, Skye, 14) or action; for example, gestures, noises, etc. *Be like* alternates with *say, think, go* and other verbs that express speech or introduce quotation.

Several scholars have assumed that discourse marker, particle and quotative *like* have all evolved out of the use of *like* as a preposition and then conjunction (Romaine & Lange 1991; Andersen 2001; Buchstaller 2001). Romaine and Lange (1991: 261) outline a path for the grammaticalisation of *like*, which very much hinges on it developing functions of the conjunction. Once *like* had reached this point, it was able to take clausal complements. *Like* introduces full sentential clauses in this function; thereby paving the way for developing discourse marker and quotative functions.

D'Arcy (2005) argues that the development of discourse marker functions is more complex than Romaine and Lange had assumed, and that the discourse marker developed from the use of *like* as a sentence adverb[2]. She further explains that the discourse particle represents the beginning of a new developmental cline in the evolution of *like*, rather than developing alongside the discourse marker. Once *like* has developed discourse marker functions, it enters syntactic structure and, with time, generalises 'from one maximal projection to another' (2005: 204). She provides

[2] Sentence adverb *like* has backward scope and occurs at the end of a proposition, for example: *We'll have to get this room cleaned up for Sam's visit* ***like***. There are not a sufficient number of tokens of this type in our data to include this function in our analysis.

evidence for her argument with apparent time data from Toronto, which suggests that clause internal *like* was a later development. Thus, the discourse particle differs from previous *like* functions in that it represents a move back into the syntax (rather than further to the edge) as well as a reduction in scope. Conversely, the discourse marker moved further towards the syntactic edge and involved scope broadening.

It is widely assumed that discourse marker *like* also predates the development of quotative *like*. The latter is believed to have emerged during the last three decades of the twentieth century (Butters 1982; Blyth *et al.* 1990; Romaine & Lange 1991; Tagliamonte & D'Arcy 2004). There is disagreement as to how precisely this happened. Several scholars have argued that *like*, in its vernacular functions, has gone through a process of grammaticalisation (Romaine & Lange 1991; D'Arcy 2005; Tagliamonte & D'Arcy 2007), that is, lexical forms have developed grammatical functions. This happens gradually and in certain contexts (Hopper & Traugott 1993: xv). According to Romaine & Lange (1991: 261), once the conjunction function was introduced and *like* was able to take clausal complements, the quotative was able to evolve in contexts where a *like*-introduced clause is a quotation. Here, *like* is reanalysed as a quotative complementiser and a dummy *be* is inserted into the quotative frame as English clauses need to contain a verb (Romaine & Lange 1991: 261).

D'Arcy (2015: 53) provides evidence that questions the grammaticalisation development outlined by Romaine & Lange, in addition to the status of *like* in *be like* as a complementiser. She cites Buchstaller's (2014) proposition as an alternative scenario by which the quotative was formed by the discourse marker filling the syntactic slot adjacent to *be*. D'Arcy (2015) also entertains the possibility that *be like* may actually be on a trajectory of lexicalisation rather than grammaticalisation, i.e. it has become a lexical rather than a grammatical structure. This relies on the assumption that the discourse marker and the verbal element *be* function as a unit: they have developed a new meaning, new constraints and new ways of use. She provides several pieces of evidence for her argument relating to a presumed lack of constraint reorganisation and context expansion (Tagliamonte & D'Arcy 2007; Durham *et al.* 2012; Haddican, Zweig & Johnson 2015; but see Butters 1982: 149; Ferrara & Bell 1995: 279; Tagliamonte & D'Arcy 2004), decategorialisation and semantic bleaching.[3]

It has been pointed out frequently that the definition of grammaticalisation and lexicalisation hinges on one's view of grammar and the lexicon. For those who do not draw a line between the two, the need to differentiate these two processes holds less importance. Vandelanotte (2012) argues that we may best view the development of *be like* as a case of constructionalisation, putting the locus of language change on the clause rather than the verb *be like*. He questions the analysis of *like* as a complementiser on several grounds (see Vandelanotte 2012: 176–9); instead analysing the construction of which *be like* forms a part as a reporting clause which functions as a conceptually dependent head and the reported clause as a conceptually autonomous

[3] In principle, of course, both grammaticalisation and lexicalisation can be involved in the development of quotatives, consecutively (Lehmann 2002) or even simultaneously (e.g. see Haas 2007 on *each other*).

complement (2012: 181). In a construction grammar framework, he argues that constructionalisation was involved in the initial formation of a *be like* construction. Through analogical extension, there was then a meaning shift from imitation clauses involving *be like*, which were brought into correspondence with reporting clauses, and were added to the inventory of reporting clauses slotting into a broader taxonomy. He views this development as the continued analogical integration of *be like* into 'the "canonical" direct speech and thought construction' (2012: 189), rejecting a grammaticalisation scenario in the strict sense that involves mechanisms such as decategorialisation as well as a scenario by which the *be like* construction may be developing into a formulaic phrase. The advantage of this analysis is that variants of the *be like* construction can be viewed as analogical extensions from *like* to particles with similar functions slotting into the same direct speech and thought construction.

This correlates with Buchstaller (2014), who treats *be like* as the combination of *be* and a discourse marker. Similarly to Vandelanotte (2012), Buchstaller (2014: 15–17) views quotation as a set of constructions that expresses reportativity and the *be like* quotative as a subconstruction of the type Noun Phrase + *be* + *like* + Quote.[4] Lexical slots in a productive construction normally can be filled with other material via analogical extension. Therefore, other words, usually discourse markers, such as *all*, *kinda*, *totally* can all fill the same lexical slot in the same way as *like*. Similarly, *be* can be replaced by other copula verbs. Thus, Buchstaller proposes a more general copula-based construction type for these innovative quotative constructions with additional schematic slots: Noun Phrase + Copula + (Discourse Marker) + Quote. Buchstaller's focus on the productivity of this construction opens up a new view to its future development. Moreover, it focuses on the emergence of a new construction, rather than solely on *be like*. We believe the focus on the reporting clause is key, and we will return to its relevance to our study in the discussion section.

While there is disagreement regarding the evolution of *be like*, it is clear that it spread into British English in the early 1990s. It was unattested until then (Tagliamonte & Hudson 1999) but has been found in several locations across England and Scotland in data gathered from the mid 1990s. It is particularly frequent among adolescent speakers but not used at all, or to very low degrees, by older speakers. Cheshire *et al.* (2011) compare different quotative functions in London. *Be like* occurs 20 to 24 per cent of the time among the younger age group in Hackney and Havering. However, when the youngest age groups are subdivided into ages 4–5, ages 8–9, ages 12–13 and ages 16–19, the use of *be like* rises to 46 per cent among the 16–19-year olds. Meyerhoff & Schleef (2013) report a similar number for native adolescents of approximately the same age in Edinburgh: 47 per cent of quotatives were occurrences of *be like*.

While previous research has focused on how the linguistic and social constraints play out in different varieties of English, system-specific pressures must surely act as an important factor influencing the phonetic form of *like* in its various functions as it

[4] Schematic positions are indicated in small caps.

spreads from one variety to another. We now turn to precisely these kinds of questions; starting with varieties outside the UK before moving on to London and Edinburgh.

### *2.2* Like *reduction*

In Drager's (2011) investigation of *like*-reduction among New Zealand adolescents, a core argument is the relationship between phonetic realisation and token frequency. It is assumed that greater reduction occurs among more frequent words and various references are cited that seem to support this assumption; for example, Bybee (2001), Zipf (1929). Furthermore, the notion is discussed that more predictable items are more likely to be phonetically reduced (e.g. Jurafsky, Bell, Gregory & Raymond 2001; Jurafsky, Bell & Girand 2002; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory & Gildea 2003). For example, Jurafsky *et al*. (2001) predict that items with a higher probability will be subject to a larger degree of reduction. Drager applies this notion to individual speakers, rather than words, as the probability that a speaker uses *like* varies between individuals. While this is an interesting notion, it should not preclude us from also giving due consideration to word predictability in our statistical models. This is particularly the case when investigating an item such as *like*, as the quotative function of it is highly limited in what lexical items can precede it: it has to be preceded by a form of *be*.

Furthermore, Drager assumes that each function of *like* is a different lemma[5] as they all have different meanings and grammatical roles. The main goal of her paper is to test whether these different lemma, which she assumes share the same word form, are realised differently phonetically. Her analysis is limited to two grammatical functions (the lexical verb and the adverb) and two discourse functions (the discourse particle and the quotative). These functions were selected as they all occur sentence-medially.

Her results reveal that the three different functions of *like* (grammatical, quotative and discourse particle) do indeed have different phonetic realisations in her data. Table 1 summarises her results. It also includes information on community-specific realisations of *like* in two communities of practice: the common-room girls and the non-common-room girls.

The largest degree of reduction is observed for the quotative, which may be surprising given frequency-based predictions and the fact that the discourse particle is reported to be more frequent than the quotative (e.g. D'Arcy 2007: 396; Drager 2011: 698). This is certainly the case for our corpora based on adolescent speech from London and Edinburgh (see table 3). Considering the much higher frequency of the discourse particle, a larger degree of reduction would be expected for this function of *like*. Many have argued for the link between word frequency and reductive sound

[5] The term *lemma* appears to be used in the psycholinguistic tradition here where a lemma represents a word's abstract conceptual form. In a two-stage model of speech production, it features in stage 1, whereas the outcome of stage 2 is the lexeme, which includes information about the pronunciation of the word. Note that this terminology differs from that used in other branches of linguistics. Here, the term *lexeme* is used for the unit of meaning. The term *stem* refers to the second stage.

Table 1. *Summary of phonetic realisations of* like *functions (adapted from Drager 2011: 703)*

| | Grammatical *like* | Quotative | Discourse particle |
|---|---|---|---|
| Pitch | Low | High | Low |
| Vowel quality | More diphthongal | Less diphthongal | Large F2 value but still diphthongal |
| /l/ to vowel duration | Long | Short | Long |
| Community of practice | Shorter /l/ to vowel duration for common room girls | More /k/ reduction for common room girls than non-common room girls | Shorter /l/ to vowel duration and more /k/ reduction for non-common room girls |

change (e.g. Bybee 2001; Phillips 2006; cf. Dinkin 2008), and although this does not refer explicitly to function frequency, a similar effect for different *like* functions is not an unrealistic expectation. However, no such link between frequency of *like* function and reduction has been documented. This finding in itself may be a clue for a contextual dependency of reduction. Indeed, Drager points towards two such potential effects: the prosodic context in which *like* occurs, particularly accentedness, and word contextual probability effects. We will test for both these factor groups in our study but will, at this point, mostly elaborate on the latter in order to clearly separate different contextual probability effects.

Drager (2011) found evidence of frequency-linked monophthongisation in New Zealand, in that speaker-predictability matters: high users reduce more. Tamminga (2013), in her study of /aɪ/-raising in *like*, did not find current evidence of the effects of *word* frequency. She compared *like* in its adjective, conjunction, discourse marker and preposition functions. With regard to /aɪ/-raising, she identifies a major divide between function and content words, rather than word frequency, in the evolution of the phenomenon.

However, there are other frequency measures, which we will now explore. The Probabilistic Reduction Hypothesis predicts that more reduction occurs in items with a higher probability (Jurafsky *et al.* 2001). The hypothesis generalises frequency-based (Zipf 1929; Rhodes 1996) and predictability-based models (Fowler & Housum 1987), as it assumes that word probability is conditioned by a whole host of contextual aspects. These may include preceding and following words, syntactic and lexical structure, discourse factors and semantic expectation. Therefore, it may be necessary to use a variety of different measures of probability to uncover probabilistic reduction effects. Jurafsky *et al.* (2001) demonstrate this for a subset of measures in their study on reduction in lexical production. While high-frequency function words were more sensitive to the predictability of neighbouring words, content words were less sensitive to the surrounding context. They demonstrated strong effects of relative frequency.

Thus, rather than focusing on one frequency measure, a combined approach appears the most logical. In this article, among other factors, we focus on one aspect of probabilistic linguistic knowledge; namely, local probabilistic relations between words. Previous research suggests that strongly related words or words that are predictable from neighbouring words are more likely to be phonologically reduced than less strongly related words (Krug 1998; Bush 1999; Bybee & Scheibman 1999). Jurafsky *et al.* (2002) argue that differences in the phonetic realisation of lemmas that have the same word form are reduced or disappear once contextual predictability is considered. If this is indeed the case and our statistical analyses reveal that it is not *like*-function but contextual factors that predict reduction, claims that items are stored in the mind in such a way that functional distinctions are maintained are much less convincing. This is precisely what Drager concludes based on her results. She argues that the different phonetic realisations for different functions of *like* give support to production models with an acoustically rich lemma level or one that is directly indexed to acoustic information.

Drager (2011: 704) argues that it is unlikely that all function-based peculiarities are due to contextual predictability, for two reasons. First, there is substantial variability in the distribution of phonetic features, and second, follow-up perception experiments showed that individuals were able to match certain *like* realisations to certain functions.

The ultimate test to determine what factors influence the form of *like* is to study the form itself while exploring as many potential predictors as possible, including *like* function. We will derive specific hypotheses based on this discussion at the end of this section. However, before we do this, we will explore what precisely the details of /l/, /aɪ/ and /k/ are in London and Edinburgh and whether there are any local system-internal factors that we should consider.

## *2.3* Like *in Edinburgh and London*

### *2.3.1 /l/*

Analysis of /l/ reduction usually focuses on lenition of the liquid consonant; in other words, whether /l/ is darkened or vocalised (Ash 1982; Carter & Local 2007; Turton 2014). Although /l/ lenition is found typically in coda position in words such as *dull* or *bulb*, rather than in onset position, as in *love* or *like*, it is not uncommon to find such lenition processes occurring in onset position in rapid or reduced speech.

In her analysis of *like*, Drager (2011) takes the duration of /l/ in relation to the following vowel in order to gauge reduction: more reduced tokens should have shorter /l/s. We will follow her in this procedure in order to render our data comparable. This procedure helps to normalise the duration of /l/. Measuring the duration of /l/ alone across different rates of speech would bias our reduction results towards fast speech. Measuring instead the duration of /l/ relative to the following vowel controls for speech rate effects.

Although most studies of London /l/ focus on vocalisation, this normally affects only coda and pre-consonantal positions. In the onset position, /l/ remains light in

Table 2. *Summary of* PRICE *realisations in London*

| RP | MC London | WC suburban London | Cockney | MLE |
|---|---|---|---|---|
| [aɪ] | [aɪ] ~ [äɪ] | [ɑɪ] | [ɒɪ] | [aː] |

London, but may be slightly palatalised (Beaken 1971: 339; Tollfree 1999: 174). Scottish English /l/, in contrast, is often described as showing dark or velarised variants in all positions (Aitken 1984: 102); however, studies conducted in Edinburgh (Speitel 1983) and Glasgow (Stuart-Smith 1999: 210) reveal that /l/ is subject to sociolinguistic variation, with middle-class and female speakers having lighter variants.

*2.3.2 /aɪ/*

Although middle-class Londoners exhibit near-RP PRICE realisations, such as [äɪ] and [aɪ], working-class suburban London English is described as having a backer nucleus, transcribed as [ɑɪ] (Wells 1982: 308; Tollfree 1999: 168; Hughes, Trudgill & Watt 2012: 77), or even rounded [ɒɪ] in more 'vigorous "dialectal" Cockney' (Wells 1982: 308).

However, this is unlikely to be relevant to the West London adolescents investigated in our study; particularly as more recent research suggests a shift from traditional Cockney forms (often associated with East London) and towards a new variety, that of Multicultural London English (MLE) (Kerswill, Torgersen & Fox 2008; Cheshire *et al.* 2011). Speakers of MLE tend to produce a more fronted and/or monophthongal [aː]-like variant (Fox 2007; Kerswill *et al.* 2008; Cheshire *et al.* 2011). Data for our corpus were collected in Ealing, a suburban district of West London,. MLE occurs less frequently in these peripheral areas of London, particularly among the white population. Nonetheless, a handful of our speakers may, impressionistically, be categorised as speakers of a near-MLE variety. However, most of our speakers produced [ɑɪ] or [aɪ] tokens. Although [ɑɪ] is still very much present in suburban London, standard [aɪ] tokens seem to be more common amongst younger white speakers (Fox 2007). PRICE realisations are summarised in table 2.

Scottish English PRICE has gained attention from being subject to the Scottish Vowel Length Rule (SVLR) in some contexts: certain vowels (particularly /i/, /u/, /aɪ/) are normally short, but lengthened before /r/, voiced fricatives, a morpheme boundary and when occurring in word-final open syllables. This means speakers have two perceptibly different diphthongs in PRICE and PRIZE, with *like* taking the shorter realisation, as it precedes a stop, giving something like [lʌik] for Standard Scottish English (SSE) speakers (Scobbie *et al.* 1999: 236). Middle-class Edinburgh speakers from areas such as Morningside may have [əɪ]~[ae] for PRICE~PRIZE, which may be neutralised to just [ae] (Chirrey 1999: 226; Stuart-Smith 2008: 58) and is often perceived as 'over-refined' (Johnston 1985: 39).

#### *2.3.3 /k/*

Variation in /k/ typically involves some kind of glottal reinforcement (preglottalisation) preceding the stop, or/and a more advanced form of lenition in full glottal replacement (glottalling). It is reported that all speakers of Edinburgh English demonstrate regular glottalisation of word-final /k/ to some extent (Chirrey 1999: 229); however, this is subject to sociolinguistic conditioning (Speitel 1983: 36). For MC speakers, the majority variant is [kʰ∼k], but WC speakers may show glottalised and fully glottalled forms [ˀk∼ʔ]. This may be a change in progress, led by young females, paralleling word-final /t/ glottalling in Edinburgh (although /k/ is some way behind).

Many Cockney speakers display frequent glottalling of the voiceless stops /p,k/ as well as /t/. This can occur intervocalically (e.g. *lucky*) and finally (e.g. *like*; Wells 1982: 323; Cruttenden 2001: 170). In Beaken's (1971: 274) study of WC children in London's East End, word-final /k/s had some form of glottalisation almost 100 per cent of the time, about half of which were fully glottalled (e.g. [ˀk∼ʔ]).

### *2.4 Conclusion*

Thus, with the exception of /k/, the phonetic detail of *like* is rather different in London and Edinburgh: *like* tokens occur in varieties with somewhat different system-internal pressures. This makes separate statistical models highly advisable. Most importantly, however, our review of /l/, /aɪ/ and /k/ indicates that all three features are variable in both cities. Some of these are already subject to reduction. This is particularly the case for /k/ in London and Edinburgh and /aɪ/ in London. Thus, if there is any evidence for function-specific reduction, we would expect it to occur in /k/ or the vowel.

Our review of the evolution of *like* would suggest that more reduction should occur in the more established vernacular functions of *like*: the discourse marker or the discourse particle. The case should be particularly strong for the discourse particle, as it is the most frequent function of *like*. However, we have also indicated in our review that contextual factors may override any of the local or function predictions. We have identified two such contextual factors to which we will pay particular attention: the prosodic context in which *like* occurs and word-contextual probability effects. We can formulate two specific hypotheses based on these deliberations:

(H0) Contextual dependency of *like* realisation: *like* function will not be a significant factor in the statistical analysis as variation is completely explained by contextual factors, e.g. prosody and word probability.

(H1) Functional dependency: *like* function will be a significant factor in the statistical analyses – alongside contextual factors, which may help us explain why function matters.

We will now turn to our methodology and explain how we will address our hypotheses and our research questions, as outlined in the introduction.

## 3 Methods

### *3.1 Data*

The data for this study come from 21 Edinburgh-born teenagers (8 males, 13 females) and 24 London-born teenagers (12 males, 12 females). The data were originally collected in the course of a study conducted on the acquisition of variation among adolescents in these locales. More detail of this can be found in Schleef, Meyerhoff & Clark (2011). Data were collected in two high schools in comparable social settings and of similar social makeup; one in West Edinburgh and one in West London. Students ranging from upper-working to lower-middle-class backgrounds were interviewed in friendship pairs in order to facilitate a casual atmosphere. The teenagers were all aged between 12 and 18, with a mean age of 14; a locally born female research assistant carried out sociolinguistic interviews in Edinburgh, and another locally born female assistant did likewise in London. The interviews were transcribed orthographically using ELAN (Brugman & Russel 2004), resulting in a time-aligned, searchable corpus.

### *3.2 Coding and acoustic analysis*

In a first step, no more than the first 55 *like* realisations per person were coded as one of the functions outlined in section 2.1 (i.e. quotative, discourse marker, discourse particle, approximative adverb, etc.). We used these for acoustic analysis. We limited these to the first 55 tokens of *like* per person in order to avoid the data from some high users of *like* biasing the analysis. We established a limit of 55, as there seemed to be a natural gap in speaker-specific token frequencies: while many speakers used up to 40 tokens of *like* in their interview, no speaker had token frequencies between 46 and 55. Conversely, if speakers did use more than 55 tokens, they would often use many more; sometimes more than 100 tokens in a single interview. This was the case for 16 speakers. For the remaining 29 speakers, all occurrences of *like* were included in the analysis as they used fewer than 55 tokens of *like* in the interview.

In addition, *all* quotatives in the transcripts were coded as one of the following: quotatives *be like*, *say*, *think*, *go*, *this is*, zero quotative and other less frequently occurring quotative verbs, such as *tell*, *shout*, etc. We used these to calculate speaker-specific probabilities to use quotative *like*. Each realisation of *like* was coded twice in ELAN by two paid, independent research assistants. Cases of coding disagreement were resolved by a third researcher.

The coded data were then extracted from ELAN and imported into a spreadsheet for further coding; for example, for preceding and following words, person, tense, speaker sex, etc. Again, this followed the same procedure of double, independent coding by two research assistants and disagreement resolution via a third party. Tokens that could not be categorised (reliably) under one of the main vernacular functions in Table 3 were coded as 'unclear/other'. This generated 804 tokens of *like* from the Edinburgh

Table 3. *Functions of* like *for adolescents in London and Edinburgh*

| | London | | Edinburgh | | Total | |
|---|---|---|---|---|---|---|
| Function | N | % | N | % | N | % |
| Discourse marker | 257 | 32.0 | 189 | 23.5 | 446 | 27.8 |
| Discourse particle | 276 | 34.4 | 323 | 40.2 | 599 | 37.3 |
| Grammatical | 154 | 19.2 | 121 | 15.1 | 275 | 17.1 |
| Quotative | 55 | 6.8 | 106 | 13.2 | 161 | 10.0 |
| Approximative adjective | 32 | 4.0 | 39 | 4.8 | 71 | 4.4 |
| Unclear/other | 29 | 3.6 | 26 | 3.2 | 55 | 3.4 |
| Total | 803 | 100 | 804 | 100 | 1607 | 100 |

speakers and 803 tokens of *like* from the London speakers. The data in Table 3 present the distribution of all functions of *like* among the teenagers recorded in Edinburgh and London. A regional comparison indicates that most functions of *like* are used in roughly the same frequency by adolescents in Edinburgh and London. The use of discourse markers is somewhat higher in London, while quotatives are used somewhat more frequently in Edinburgh. In both locations, the discourse particle is the most frequent function of vernacular *like*.

The first 55 *like* tokens per speaker were then subjected to an acoustic analysis. Firstly, the vowel and the preceding and following element, /l/ and /k/, were determined and then labelled in Praat (Boersma 2001). To determine the boundaries between sounds, we followed Drager (2011: 698) as closely as possible. Vowel formants in 10 ms intervals, duration of /l/, vowel and /k/were then measured using a Praat script. Speech rate was also determined by means of a Praat script, which calculates vowels per second based on a three-word window in Praat; that is, three words either side of each *like* token.

The preceding and following phonological environment of each *like* and the quality of the /k/ were also determined, following Drager (2011: 698) as closely as possible. For the latter, the categories were as follows: /k/ is (a) present; (b) deleted; (c) reduced but with release; (d) reduced but there is no evidence of release in the spectrogram; (e) realised as a fricative; and (f) glottalled.[6]

In the next step, prosodic characteristics were determined. The two words before *like*, and the two words following *like* (if any) were noted and the pitch pattern inspected while listening to the utterance. It was noted (1) whether *like* is accented or unaccented, and (2) whether *like* is (intonation) phrase-final or not phrase-final. Finally, (3) the boundary strength following *like*, using the Break Index Tier (Beckman &

[6] We coded for three subcategories here: (a) full-glottal: /k/ is not present and there is clear evidence of closure and release burst in the spectrogram; (b) creaky voice: /k/ is not present and there is creak rather than a glottal stop; (c) mid-glottal: /k/ is not present. This category was selected when assignment to one of the other categories was not completely certain.

Hirschberg 1993) was determined. A break index is a numerical value that is meant to represent perceived degrees of juncture in an utterance. Break indices can be assigned to perceived junctures between words and between the final word and the silence at the end of the utterance. Ratings range from 0 to 4. Beckman & Hirschberg (1993: 1–2) outline the break index values as follows:

0 for cases of clear phonetic marks of clitic groups; e.g. the medial affricate in contractions of 'did you' or a flap as in 'got it'.
1 Most phrase-medial word boundaries.
2 A strong disjuncture marked by a pause or virtual pause, but with no tonal marks; i.e. a well-formed tune continues across the juncture. OR a disjuncture that is weaker than expected at what is tonally a clear intermediate or full intonation phrase boundary.
3 Intermediate intonation phrase boundary; i.e. marked by a single phrase tone affecting the region from the last pitch accent to the boundary.
4 Full intonation phrase boundary; i.e. marked by a final boundary tone after the last phrase tone.

Strictly speaking, this is not an ordered factor as perceived juncture and intonation combine to result in a score; however, the intonation phrase does not enter as a criterion until break index 2. The break index can provide an indication of the prosodic position in which *like* occurs. Specifically, it will be an indication of the perceived break following *like* and where in the intonation phrase different functions of *like* may occur. Position within the intonation phrase and the type of break following *like* may influence the phonetic detail of *like*. We would expect *like* tokens occurring in a 0 or 1 environment to be more reduced than those occurring in a break index 3 environment.

Before processing the formant values, the database was inspected for errors and formant traces were corrected by hand. The F2–F1 distance was calculated, and a line was then fitted to the values for the F2–F1 distance. The value of this slope is used as the measure of monophthongisation, referred to as the DIPH value. If the result of the slope is positive, the space between F1 and F2 is increasing; thereby indicating a particular token is more diphthongal. A value closer to zero indicates that the vowel quality is the same throughout, and therefore more monophthongal. Due to the focus on the F2–F1 distance, there was no need to normalise the data.

Moreover, we included various measures of probabilistic linguistic knowledge. Following Drager (2011: 697), we calculated the speaker-specific relative token frequency, or probability of the quotative based on all occurrences of quotatives in the data. The measure was calculated by dividing the number of times a speaker produces quotative *like* by their total number of quotatives. Speaker-specific probabilities for other functions of *like* were not possible, given the difficulty in defining the envelope of variation for these. We also extracted bigram frequency information for our data using the SUBTLEX UK corpus (Heuven, Mandera, Keuleers & Brysbaert 2014), which consists of 201.3 million words. The SUBTLEX corpora, based on film and TV subtitles from BBC broadcasts, are new and improved frequency measures for US and UK English. SUBTLEX contains data from over 45,000 broadcasts. Therefore, they include actual speech alongside scripted speech. While this offers an advantage over corpora based on written texts (e.g. Brown, CELEX) and much smaller speech corpora (e.g. the spoken component of the BNC), we approach the results with some caution, given that scripted speech has been shown not to reflect the proper linguistic and social conditioning of variables such as *be like* (Dion & Poplack 2007).

Nevertheless, until there is a corpus of natural speech large enough to cope with the statistical measures used in the present article, SUBTLEX is by far the most appropriate tool for bigram and frequency analyses. Figure 3 illustrates the good degree of overlap in bigrams between our corpus and SUBTLEX. The presence of bigrams, such as *was like* in SUBTLEX, also demonstrates its reflection of casual speech. In addition, we are assured in our choice by work in psychology which has demonstrated that word processing times are highly improved in SUBTLEX, compared with previously used frequency estimates (see Brysbaert & New 2009).[7]

SUBTLEX bigram frequencies are further used as a measure of conditional probability (Jurafsky *et al.* 2001), by calculating how likely a word is to occur given the previous word was *like*, and how likely *like* is to occur given the previous word. The conditional probability is derived by counting how often the two words in question occur together and then dividing this number by the number of times the first word occurs. Independent of the measure we take, we would expect more reduction with higher probabilities.

In an attempt to make our results comparable with those of Drager (2011), we focused on specific functions of *like* in our analysis. We conflated grammatical functions of *like* into one category; specifically, lexical verb and adverb were conflated. These were contrasted with other functions of *like*: the quotative, the discourse particle and the discourse marker. Approximative adjectives do not form part of the analysis as token counts were deemed too low to yield any reliable results. Unclear and 'other' tokens (such as utterance-final sentence adverbs) were also removed from analysis. The independent variables used in the current analysis, and the levels coded for under each variable, are outlined below:

- Variant of like: grammatical, discourse marker, discourse particle, quotative *be like*
- /l/ to V ratio between duration of /l/ and vowel: continuous (higher values indicate a longer /l/ to vowel duration)
- Speech rate based on vowels per second: continuous
- Duration of /l/, vowel and /k/ in milliseconds (ms): continuous
- Bigram predictability – bigram count 1: (*like* + word), bigram count 2: (word + *like*): continuous
- Conditional probability as above, but with conditional probabilities instead of counts: continuous
- Speaker-specific probability of using quotative *like*: continuous
- Diphthongisation (DIPH) value – high values indicate more diphthongal realisation of vowel: continuous
- /k/ realisation: present, deleted, reduced, fricated, glottalled
- Accentedness of *like*: accented, unaccented

[7] The downside of using bigrams for this dataset is that they exclude sentences which begin or end with *like* from the relevant measure.

- Position in intonation phrase: *like* is phrase-final, *like* is not phrase-final
- Boundary strength (BS) following *like*: 0, 1, 2, 3, 4[8]
- Preceding phonological context: vowel, pause, nasal, liquid, plosive, fricative and affricate, glide
- Following phonological context: vowel, pause, plosive, consonant other than plosive[9]
- Speaker sex: male, female

## 4 Results

Results are given in four distinct steps. In the initial step, we provide a summary characterisation of the different functions of *like*. We then explore the vowel quality of *like* and finally proceed to investigate /l/ and /k/.

### *4.1* Like *function as dependent variable*

Drager (2011) focuses on predicting the particular variant of *like* by using the probability of the variant as the response variable and the phonetic and contextual values as the predictors. This is somewhat unusual, given that the dependent variable (i.e. the variable being tested) typically would be the phonetic realisation, and the discourse function would be the predictor. We run analyses of this kind in section 4.2; however, in order to provide a comparison with Drager (2011), section 4.1 uses the probability of *like* variant as the dependent variable, to assess whether a particular variant can be predicted from its phonetic realisation.

Our data differ from Drager's in some ways, and we believe we can improve on exploring the complexity of our data by using different statistical techniques. Drager considers three functions of *like* in her paper, resulting in three models. However, as we have four possible variants – discourse marker (DM), discourse particle (DP), grammatical (Gram) and quotative (Quot) – we will need six models to be able to consider each pairing. Thus, we are asking whether a particular variant of *like* (e.g. whether it is a discourse marker or a quotative) can be predicted by the phonetic realisation of each of its segments, as well as other contextual factors listed in section 3.2. If indeed the phonetic factors emerge as the most important, this would lend weight to the idea that items are stored in the mind in such a way that functional distinctions are maintained.

Because of the large number of predictors for this model, we took advantage of the binary choice in dependent variables by running a series of random forests. This has the added benefit of including highly collinear predictors (such as bigram counts and their conditional probability, or accentedness and boundary strength) in the same model, as random forests are well equipped to deal with such predictors, as well as small sample sizes.

[8] We excluded boundary strength 4 tokens, due to their low token numbers.

[9] Originally coded for plosive; fricative and affricate; glide; nasal and liquid, but conflated on the basis of similar effects in regression models.

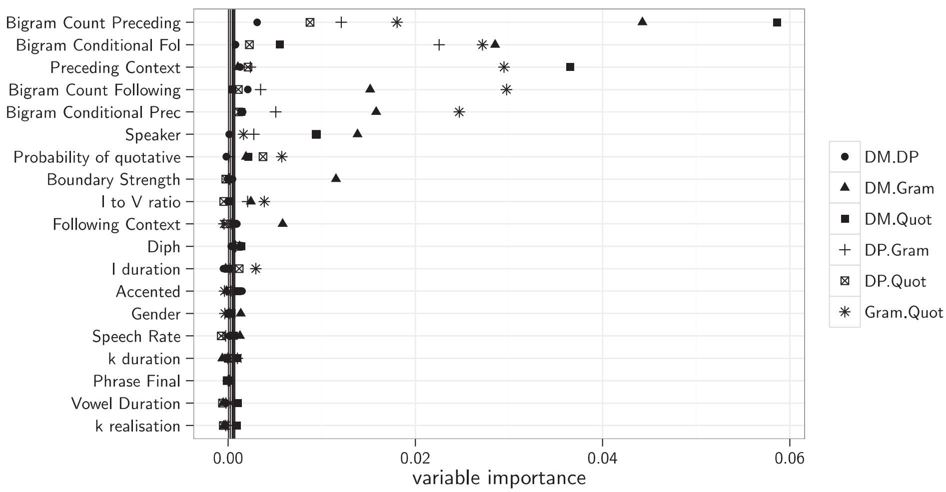


Figure 1. Variable importance of factors from random forest in Edinburgh

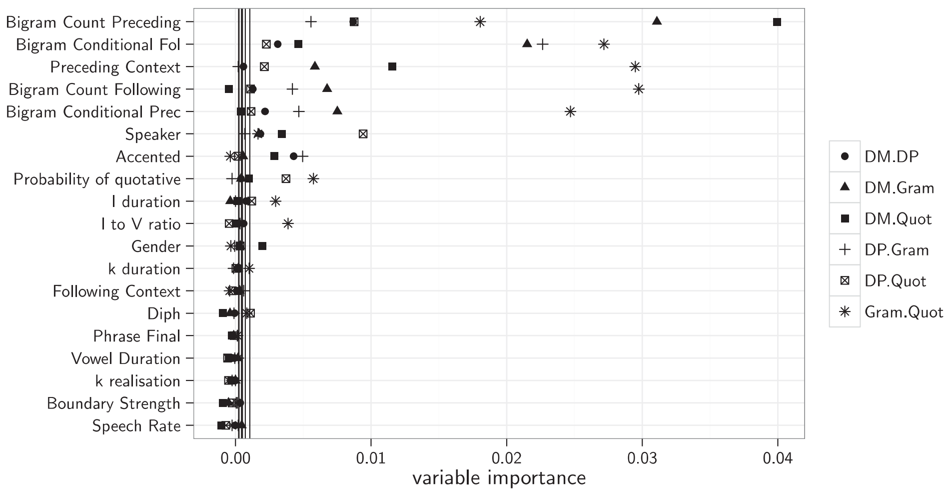


Figure 2. Variable importance of factors from random forest of London

The importance of the predictors for Edinburgh and London can be viewed from the dotplots in figures 1 and 2, alongside the ctrees in the Appendix. Each shape represents a different pairwise combination of variants of the dependent variable. The lines represent the cut-off for significance: everything to the left of the line is considered non-significant in each comparison (Strobl, Malley & Tutz 2009: 342). Although there are six separate lines on each of these plots, they are very tightly clustered, and any points approaching the lines should be taken as not significant.

Overall, we can see that bigram frequency of the preceding and following words, and the preceding context overwhelmingly are the most important factors for predicting

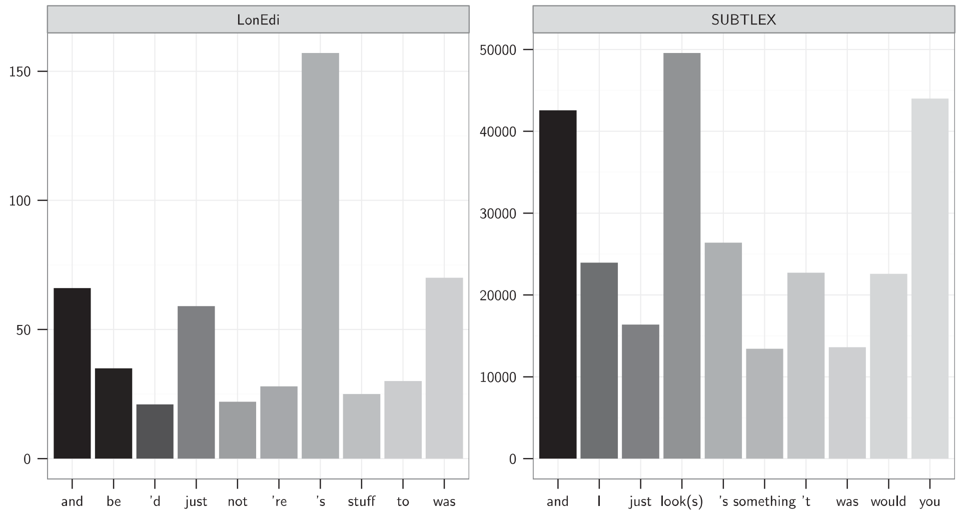


Figure 3. Top ten bigram frequencies for like and the preceding word in our corpus and in the SUBTLEX corpus

what kind of variant of *like* we have.[10] For example, frequent combinations with *like* and a specific preceding word (BigramCount2) relate to a variety of functions of *like*, not one specific function. Depending on the comparison, discourse markers and discourse particles are often associated with low bigram frequencies of that particular type, while the grammatical and quotative functions are often associated with high bigram frequencies. Some frequent combinations are listed in figure 3.

Boundary strength, accentedness, /l/ and /l/ to vowel duration, and following context also play a role for most or some combinations in both cities. In combinations that include the quotative, we also find the probability of a speaker using the quotative to be a significant predictor. This finding is unsurprising and somewhat circular: if a token of *like* is uttered by a speaker who has a high probability of using quotative *like*, then that token is more likely to be quotative *like* in comparison to other speakers.

The ctrees, built using R's party package (Hothorn, Hornik & Zeileis 2006), give us more of an insight into how each factor contributes. As an example, figure 4 illustrates a conditional inference tree for discourse marker vs grammatical *like* in Edinburgh. If the bigram with the following word is not particularly frequent (less than 3,350 occurrences in SUBTLEX), it is more likely to be a discourse marker. If it is very frequent, the likelihood of the variant is decided by the boundary strength: a weaker boundary indicates a grammatical element, whereas a stronger boundary is more likely to be preceded by a discourse marker. The other trees (which can be viewed in the Appendix) vary in their complexity; however, they demonstrate that the variant can be accounted for with similar patterns of contextual factors.

[10] SUBTLEX classifies a preceding contracted verb form as the previous word.

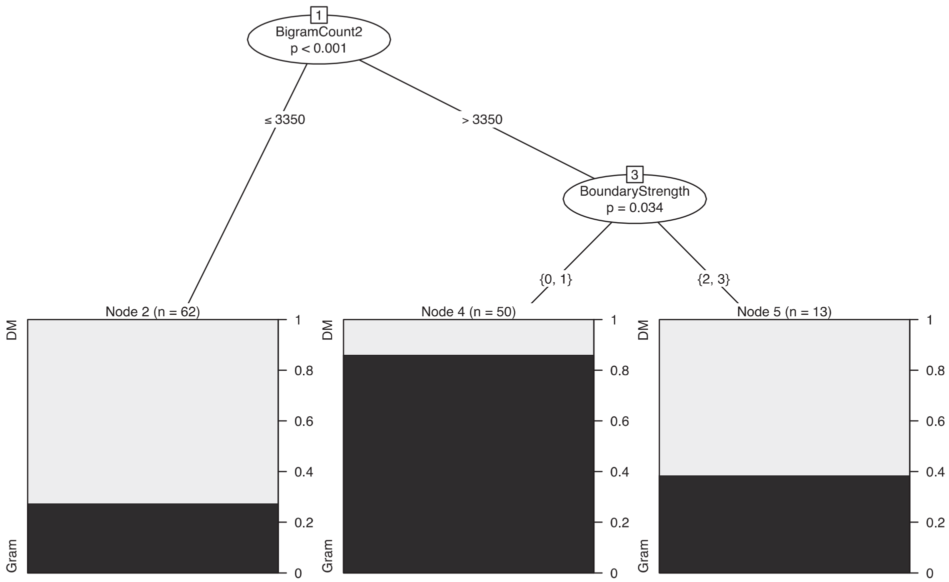


Figure 4. Conditional inference tree for discourse marker (DM) vs grammatical (gram) *like* in Edinburgh

The forests and trees for this dataset inform us that the main distributional facts reflect the fact that the context of the sentence (i.e. bigrams or boundary strength) is the overwhelming predictor of the type of *like*.[11] However, there are some small indications, in some combinations that contain the quotative, that phonetic factors may play a role in determining the type of *like*. Diphthongisation/monopthongisation, /k/ and /l/ to vowel duration demonstrate some small effects.

We will now explore more closely these phonetic factors and create three models that take the phonetic realisation as the dependent variable; that is, the l-to-vowel duration, the degree of monophthongisation/diphthongisation and the quality of /k/ respectively. This tests whether different functions of *like* do nonetheless differ significantly in the quality of /l/, /aɪ/ and /k/, despite the contextual constraints. Therefore, the research question in this section is somewhat different. Rather than asking whether a particular variant of *like* (e.g. whether it is a discourse marker or a quotative) can be predicted by the phonetic realisation of each of its segments, as well as other contextual factors, we are now asking whether monophthongisation, /k/ quality and l-to-vowel duration can be predicted from the factors listed in section 3.3. The

[11] To test whether the overwhelming strength of the bigram counts and conditional probabilities might be obscuring fine-grained phonetic effects, we conducted a random forest analysis without the bigram figures and preceding and following context. We find very similar variable importance results for these trees: individual speaker differences become the strongest predictor and boundary strength and accentedness rank highly in many combinations.

Table 4. *Coefficients of a general mixed-effects linear regression of monophthongisation and diphthongisation with speaker as random effect – London*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std error | t-value | p-value |
| (Intercept) | 0.643 | 0.067 | 9.547 | <0.001 |
| Boundary strength (ref. level: 0) | | | | |
| Boundary strength 1 | – 0.190 | 0.055 | – 3.451 | <0.001 |
| Boundary strength 2 | – 0.147 | 0.061 | – 2.429 | 0.015 |
| Boundary strength 3 | – 0.168 | 0.057 | – 2.930 | 0.004 |
| Variant (ref. level: discourse marker) | | | | |
| Discourse particle | – 0.001 | 0.021 | – 0.053 | 0.958 |
| Grammatical | – 0.044 | 0.025 | – 1.777 | 0.076 |
| Quotative | – 0.082 | 0.034 | – 2.425 | 0.016 |
| Standard deviation: 0.221 | | | | |

most important question here is: will the function of *like* be selected as a significant predictor or not?

### 4.3 Vowel quality

Tables 4 and 6 provide the results of the linear regression models for the vowel quality in London and Edinburgh. They list significant factors in column 1, and in columns 2, 3, 4 and 5, the estimate, the standard error, the t-value and the p-value are listed respectively. We further provide the estimate of the intercept. The estimates given in the regression models help us interpret the effect of a relevant factor level on vowel quality. A negative quantity implies a larger degree of monophthongisation, while a positive quantity implies a larger degree of diphthongisation than the respective reference level.

We can see from table 4 that among our London adolescents, two factors achieve statistical significance. *Like* tokens followed by a boundary strength of 1 are the most monophthongal *like* tokens, followed by boundary strength 3.[12] Thus, prosodic factors do indeed influence the realisation of *like*, as we had expected. In addition, the function of *like* does influence the degree of monophthongisation as well: quotatives are more monophthongal among London adolescents, which mirrors Drager's results in New Zealand.

The box plot in figure 5 summarises the results for variants in London. The additional bars demonstrate that quotatives are significantly more monophthongal than discourse markers and discourse particles. Discourse particles differ significantly from quotative and grammatical *like*, and grammatical functions of *like* are only

[12] The difference between 0 and all other junctures is significant. *Like* tokens preceding a break strength of 0 are more diphthongal than *like* tokens preceding other break types. The difference between 1, 2 and 3 is not significant, which was confirmed by switching the default intercept factor group in turn.

Table 5. *Occurrence of* like *in five boundary strength contexts in both London and Edinburgh*

| | **0** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|---|
| Discourse marker | 3 | 209 | 54 | 173 | 4 |
| Discourse particle | 29 | 379 | 58 | 128 | 1 |
| Grammatical | 18 | 210 | 13 | 29 | 2 |
| Quotative | 1 | 113 | 13 | 31 | 0 |

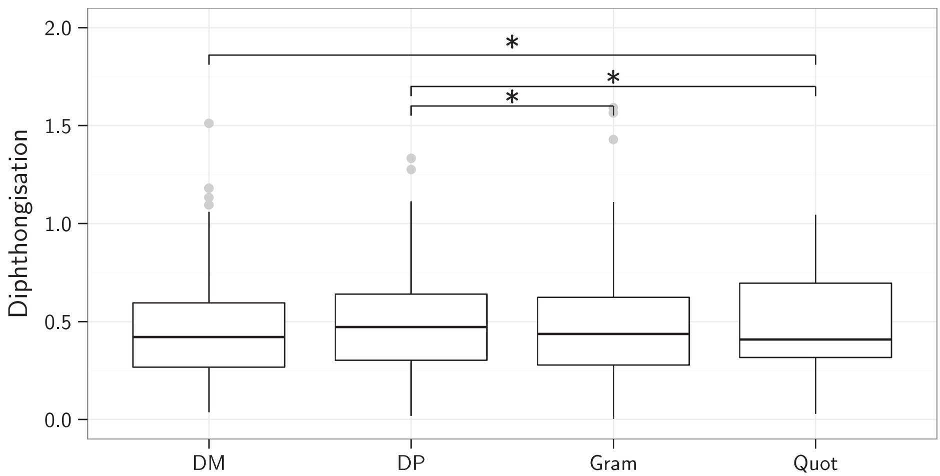


Figure 5. Degrees of diphthongisation of *like* in its different functions – London

significantly different from discourse particles. Although significant, the boxplot shows that the differences are very fine-grained. Moreover, the results point towards the conclusion that, in London, break type 1 is a favourable environment in which monophthongisation is more frequent. Table 5 outlines the breakdown of break type by *like* function.

Table 6 shows that the results for Edinburgh are very similar to our London findings. Boundary strength (BS) is a significant factor. However, in Edinburgh, *like* tokens followed by a juncture strength of 3 are the most monphthongal *like* tokens, followed by BS 1. In contrast to BS2, BS1 and BS3 are less diphthongal than BS0, although this is only marginally the case for BS1. The difference between 1, 2 and 3 is not significant. We will return to this point in our discussion.

As in London, the function of *like* influences the degree of monophthongisation. Figure 6 shows that discourse particles are most diphthongal – more so than discourse markers, and significantly more so than grammatical *like* and quotatives. Discourse markers are the next most diphthongal, but not significantly more so than grammatical *like*. Quotatives are the most monophthongal, with significantly lower DIPH values than discourse particles and discourse markers, but not grammatical *like*.

Table 6. *Coefficients of a general mixed-effects linear regression of monophthongisation and diphthongisation with speaker as random effect – Edinburgh*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std error | t-value | p-value |
| (Intercept) | 0.848 | 0.077 | 11.007 | <0.001 |
| Gender (ref. level: female) | – 0.197 | 0.056 | – 3.484 | 0.002 |
| Boundary strength (ref. level: 0) | | | | |
| Boundary strength 1 | – 0.099 | 0.058 | – 1.691 | 0.091 |
| Boundary strength 2 | – 0.059 | 0.069 | – 0.847 | 0.397 |
| Boundary strength 3 | – 0.136 | 0.064 | – 2.141 | 0.033 |
| Variant (ref. level: discourse marker) | | | | |
| Discourse particle | 0.049 | 0.031 | 1.592 | 0.112 |
| Grammatical | – 0.065 | 0.040 | – 1.620 | 0.106 |
| Quotative | – 0.108 | 0.042 | – 2.568 | 0.010 |
| Vowel duration | – 2.035 | 0.505 | – 4.031 | <0.001 |
| Standard deviation: 0.314 | | | | |

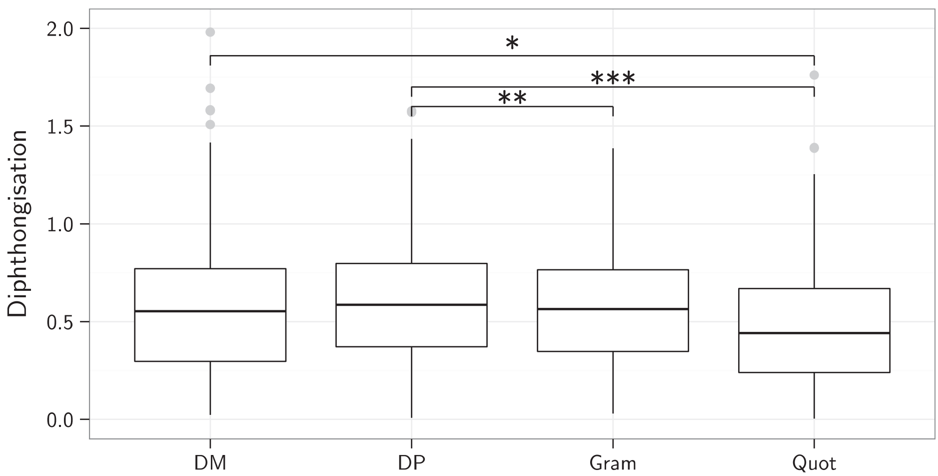


Figure 6. Degrees of diphthongisation of *like* in its different functions – Edinburgh

In Edinburgh, there is also a gender effect: males are more likely to realise the vowel as a monophthong across all four functions. There is no significant interaction effect between gender and variant; thereby indicating that the gender effect occurs independently of *like* function. The fact that males are more likely to realise the vowel as a monophthong is not terribly surprising. Many studies have shown that men use vernacular features more than women, especially in cases of stable variation and language change above the level of awareness (Labov 2001). However, given

Table 7. *Coefficients of a general mixed-effects linear regression of /l/ to vowel duration with speaker as random effect – London*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std error | t-value | p-value |
| (Intercept) | 0.510 | 0.053 | 9.6040 | <0.001 |
| Boundary strength (ref. level: 0) | | | | |
| Boundary strength 1 | – 0.047 | 0.046 | – 1.030 | 0.303 |
| Boundary strength 2 | – 0.079 | 0.050 | – 1.576 | 0.115 |
| Boundary strength 3 | – 0.096 | 0.047 | – 2.029 | 0.042 |
| Gender (ref. level: female) | 0.099 | 0.041 | 2.379 | 0.026 |
| Standard deviation: 0.094 | | | | |

the relatively recent introduction of *like* and its use by younger age groups, we are not in a position to make an assessment about the nature of the variation; that is, whether or not it, or even the vowel quality of /ai/, is stable. It is similarly unlikely that monophthongised *like* occurs above the level of awareness (yet), given the recentness of the phenomenon. Future research is needed to uncover what precisely the status of the observed variation is: by collecting data from different age groups and by conducting perception and ethnographic work to investigate the extent to which monophthongisation is perceived by speakers and the stylistic work it may do.

The significant factor of vowel duration indicates that long vowels are more monophthongal. The situation is somewhat more complex but space constraints do not permit a full treatment of this issue. In brief, more fine-grained analyses suggest by selecting a binary cut-off between monophthongal (i.e. a DIPH value of around 0) and diphthongal tokens, duration correlates with the diphthongal tokens only.

### 4.3 /l/ to vowel duration

The /l/ to vowel duration provides an indication of the length of /l/ relative to the vowel: a low /l/ to vowel ratio indicates a short /l/ relative to the vowel duration. More reduced tokens should have shorter /l/s.[13] In Drager's (2011) study, tokens with a short /l/ to vowel ratio were more likely to be quotative *like* than grammatical *like*. The results for /l/ to vowel duration in London (table 7) and Edinburgh (table 8) indicate that *like*-function does not correlate significantly with the duration of /l/. Considering this finding, we provide only a brief report of the results.

[13] We note that existing analyses of /l/ reduction have shown that darker /l/s have a longer duration (e.g. Sproat & Fujimura 1993). However, this may be true only of categorically dark variants, i.e. not initial /l/ (Yuan & Liberman 2009, 2011). A somewhat more reliable way to determine /l/ reduction in future work may be to take the acoustic correlate of darkness i.e. the distance between F2 and F1: the lower the value, the darker the /l/.

Table 8. *Coefficients of a general mixed-effects linear regression of /l/ to vowel duration with speaker as random effect - Edinburgh*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std error | t-value | p-value |
| (Intercept) | 0.579 | 0.020 | 28.743 | <0.001 |
| Preceding context (ref. level: fricative) | | | | |
| Preceding stop | 0.058 | 0.022 | 2.656 | 0.008 |
| Preceding pause | 0.014 | 0.022 | 0.650 | 0.515 |
| Preceding vowel | 0.080 | 0.020 | 3.953 | <0.001 |
| Standard deviation: 0.071 | | | | |

In London, the boundary strength following *like* has a significant effect. A BS of 0 is associated with a longer /l/ to vowel duration, while BS1 and 3 reflect the results we found for vowel quality in that they are associated with a larger degree of reduction – and in relation to /l/ – a shorter /l/ to vowel duration. Table 7 also indicates that males have a higher /l/ to vowel duration.

In Edinburgh, conversely, the preceding phonological context constrains /l/ to vowel duration in *like*. The /l/ to vowel duration is highest when the preceding context is a vowel, followed by stops, pauses and fricatives. The fricative context, which results in the shortest /l/ to vowel duration, differs significantly from vowels and stops, but not pauses.

### *4.4 /k/ realisation*

The results for /k/ in London (table 9) and Edinburgh (table 10) also indicate that *like* function is not a significant factor in determining the realisation of /k/. In tables 9 and 10, a higher value indicates an increased likelihood of /k/ reduction.

In London, again the boundary strength following *like* has a significant effect. A BS of 0 is associated with the fullest /k/ realisation, while BS1 is significantly more reduced. Table 9 also indicates that fuller /k/ realisations are associated with more diphthongal realisations of the vowel, suggesting that /k/ and vowels in *like* are reduced in tandem. Finally, the realisation of /k/ is influenced significantly by the following context. This is the case in London as well as Edinburgh (see table 10). In both locales, we observe the highest degree of /k/ reduction before stops and the least degree of reduction before vowels. Other consonants and pauses fall in between these two extremes. This is in line with findings for /t/ reduction, particularly T-glottalling, for which a consonant > pause > vowel hierarchy is often documented (see Schleef 2013 for an overview).

We can also see that in Edinburgh the number of times *like* and the preceding word occur together shows a significant effect. Figure 7 indicates that a fully realised /k/ is

Table 9. *Coefficients of a general mixed-effects linear regression of /k/ with speaker as random effect – London*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std error | z-value | p-value |
| (Intercept) | 2.073 | 1.078 | 1.922 | 0.055 |
| Following context (ref level: stop) | | | | |
| Following consonant (other than stop) | – 3.256 | 0.570 | – 5.713 | <0.001 |
| Following vowel | – 3.712 | 0.588 | – 6.309 | <0.001 |
| Following pause | – 3.160 | 0.690 | – 4.579 | <0.001 |
| Boundary strength (ref. level: 0) | | | | |
| Boundary strength 1 | 2.580 | 0.966 | 2.669 | 0.008 |
| Boundary strength 2 | 1.835 | 1.038 | 1.768 | 0.077 |
| Boundary strength 3 | 1.737 | 1.059 | 1.640 | 0.101 |
| Diph | – 1.924 | 0.414 | – 4.646 | <0.001 |
| Standard deviation: 0.969 | | | | |

Table 10. *Coefficients of a general mixed-effects linear regression of /k/ with speaker as random effect – Edinburgh*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std error | z-value | p-value |
| (Intercept) | 1.522 | 0.488 | 3.114 | 0.001 |
| Following context (ref level: stop) | | | | |
| Following consonant (other than stop) | – 2.508 | 0.414 | – 6.058 | <0.001 |
| Following vowel | – 3.172 | 0.428 | – 7.408 | <0.001 |
| Following pause | – 2.612 | 0.404 | – 6.460 | <0.001 |
| Bigram Count 2 (log scaled) | 0.105 | 0.040 | 2.592 | 0.009 |
| Standard deviation: 0.686 | | | | |

the most common variant found, but if the bigram is very frequent it is more likely to be reduced in some way (log-scale between 9 and 10; that is, between about 8,000 and 22,000 occurrences in the SUBTLEX corpus). This is in accordance with the predictions made by Jurafsky *et al.* (2001, 2002).

In conclusion, both results for /l/ and /k/ confirm our findings made in the random forests analysis: the function of *like* correlates mostly with contextual factors rather than phonetic factors. Predictability of the preceding and following words, the preceding and following segments, and BS emerge as highly significant. Different

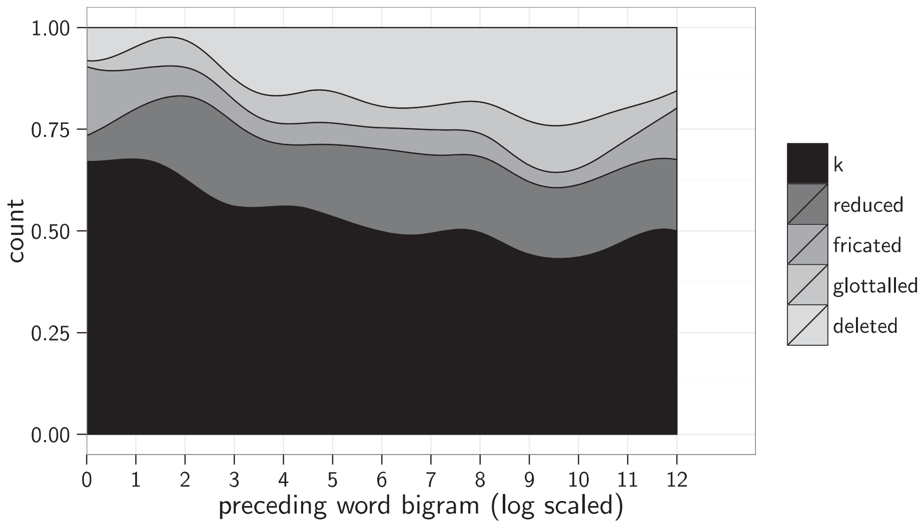


Figure 7. /k/ realisation for Edinburgh speakers against the log of bigram count 2 (i.e. number of times like and its preceding word occur together)

functions of *like* occur in specific contexts, which have not resulted in significantly different phonetic details for them.

However, this does not explain all the findings pertaining to vowel realisation. In London and Edinburgh, the vowel is the only non-contextual feature that seems to be sensitive to the function of *like*: quotative *be like* is more likely to be monophthongised than other functions of *like*. This means there are three independent speech communities (New Zealand, London and Edinburgh) across which the PRICE vowel is subject to different phonetic and phonological pressures, and in all three the data show that the quotative is most monophthongal. This is highly suggestive of a contextual explanation for the monopthongisation of *like* in this function.

## 5 Discussion

### *5.1 Summary*

The discussion focuses on the monopthongisation of quotative *like*. It attempts to provide an explanation for this phenomenon and illuminate its relevance. We argue that it is rather premature to make claims about the cognitive representation of functions of *like* based on our evidence and that the more monophthongal nature of quotative *like* is due to the context in which it occurs. Before we develop a clearer notion of the contextual position in which *be like* has developed in section 5.3, we summarise and explain the relevance of the findings we've made for our contextual factors and our frequency factors in section 5.1, and we discuss the potential relevance of our findings to theoretical arguments regarding the evolution of *be like* in section 5.2.

Our results reveal that *like* function is an impressive predictor of monophthongisation, yet never on its own: contextual factors appear to matter as well. Our data suggest that quotative *be like* is most frequently followed by a BS1 break (see table 5); in other words, it is not followed by a strong disjuncture and occurs within, rather than at the edge of, an intonation phrase boundary: quotative and quote tend to be part of the same intonation phrase. The other three *like* types also occur frequently in this prosodic context; however, in contrast to *be like*, they also occur quite frequently in other prosodic contexts (see table 5). Those in which *be like* occurs are less variable than those of the other three *like* functions. While the grammatical function of *like* is similarly restricted to *be like* with regard to its prosodic context, in contrast to *be like*, it also occurs in a BS0 context. Thus, prosody does seem to play an important role in the contextual embedding of different functions of *like*.

What makes these findings relevant is that the monopthongisation of /aɪ/ is constrained in a similar manner to the prosodic embedding of *be like*: quotative *like*, more often than other *like* functions, occurs in a favourable environment for monophthongisation. The most diphthongal tokens of *like* occur in a BS0 environment, which is a context in which *be like* hardly ever occurs. In London, *like* tokens in BS1, 2 and 3 contexts are more likely to be more monophthongal. They represent favourable environments for monophthongisation. Thus, *like*, here, appears in a slot where reduction is likely to occur. This makes quotative *like* tokens least likely to be diphthongal, as they seldom occur in a BS0 context.

In Edinburgh, the situation is similar, yet more complex: the most diphthongal tokens of *like* also occur in a BS0 environment, which is a context in which *be like* hardly ever surfaces. Here, *like* tokens in BS3 contexts are more likely to be monophthongal, the second most frequent context for *be like* to occur. This situation is rendered more complex by the fact that BS1 and BS2 contexts do not differ significantly from BS0 contexts, BS3 contexts and each other: they lie in between these two extremes. Nonetheless, quotative *like* tokens are least likely to be diphthongal, as they seldom occur in a BS0 context; moreover, most *like* tokens occur in an environment in which monophthongisation can certainly happen. Thus, in both locations *like* appears in a slot where reduction is likely to occur. The fact that in Edinburgh the situation is more complicated and that in both locations BS1, BS2 and BS3 do not differ from each other at a significant level, suggests that BS is merely a reflection of a more important underlying process, and that our study may not have caught fully the factors that constrain monophthongisation. Something else seems to be going on. We will return to this below.

While prosody does play a role, bigram frequency does not seem to be a contextual factor that influences the monophthongisation of *like*, as we had anticipated. Our models of variable importance tell us that the preceding word does matter (in the prediction of *like*, not a specific function of it); however, our analysis of monophthongisation, where function emerges as a significant effect, indicates that bigram *frequency* by itself does not interact with *like* function. For example, while *was like* is a frequent bigram that predicts the quotative, other highly frequent bigrams predict other functions; for example, *I like* is grammatical, *and like* tends to be a discourse marker or discourse particle.

### *5.2 Development of be* like

When we consider the proposals regarding the process through which quotative *be like* has developed – grammaticalisation, lexicalisation and constructionalisation – our conclusions have to remain rather modest. Our findings are only tangentially relevant to this debate: they confirm that one of the *like* functions, quotative *be like*, has been placed in a position to split off in form from the others. In section 5.3, we offer an explanation of what exactly this position is. Since our data are not of a historical nature, we have little to contribute to the discussion of how *like* got into this position because we would expect reduction in all three processes implicated.

Grammaticalisation normally includes (a) pragmatic shift, (b) desemanticisation, (c) decategorialisation and (d) phonetic reduction (e.g. Bybee 2003). This makes sense, as grammaticalised items tend to be used more frequently and hence become more predictable, thereby resulting in the loss of phonetic substance (Heine 2003: 583). Highly grammatical elements, such as auxiliaries and prepositions, are often subject to reduction. Grammatical elements tend to be more reduced than lexical ones (e.g. nouns).

It is clear from our data that reduction is associated more strongly with the quotative than it is with the other *like* functions. However, the findings of our study do not mean that *be like* has been grammaticalised because both grammaticalisation and lexicalisation may involve reduction (Lehman 2002: 1). Lexicalisation is in fact very strongly associated with reduction (Brinton & Traugott 2005). An analysis of *be like* as part of a construction too would lead us to expect reduction because reporting clause constructions generally tend not to be particularly prominent prosodically (e.g. Halliday & Matthiessen 2004: 446). We believe prosodic constraints are key to understanding why it is that *like* in the quotative is reduced more than the other functions of *like*. The remainder of our discussion focuses on this issue of reduction rather than the evolution of *be like*, and we develop a clearer notion of what precisely it is that reduces *like* in the *be like* quotative more that the *like* serving other functions.

### *5.3* Like *reduction*

We would like to propose that our findings for monopthongisation and break type are due to one unifying development. The evolution of *be like*, by whatever process, has brought quotative *like* under the influence of rhythmic and reduction patterns of reporting clauses which are fundamentally an outcome of their relation to the reported clause.

Halliday & Matthiessen (2004: 446) argue that in spoken English, the reporting clause is normally less prominent than the reported clause. It is proclitic, i.e. – in Halliday's terms – non-salient and pre-rhythmic, if it comes first. There is a very good reason for this: the reporting clause has only one function; namely, to indicate that the reported clause is projected.

Clauses such as *he said*, *I was like* and *she's like* have one thing in common: they usually consist of precisely one foot (or rhythm group). In English, each foot consists of one or more syllables. Some syllables carry stress, while others do not. These are often referred to as strong/salient and weak respectively (Halliday & Matthiessen 2004: 12). In English *natural speech* there is a tendency for strong syllables to occur in equal intervals, and stress may shift in order to maintain an even alternation of strong and weak beats. In verbs, the final syllable is said to carry main stress if it is heavy, which it is in *be like*. However, sentential stress tends to shift as the first word of the reported clause is likely to be stressed. Thus, *like* is often weakened because it occurs right before the beginning of the next foot, the beginning of the reported clause.[14]

Indeed, our own data show this. Quotative *like* tends not to be accented. While stress and accent are not the same thing, the fact that quotative like is unlikely to receive an accent is certainly also due to the argument presented above: its position in the reporting clause. Only 13 per cent of *like* tokens in quotatives are accented, in contrast to 44 per cent of discourse marker *like*, 17 per cent of discourse particle *like* and 37 per cent of grammatical *like* functions. Nonetheless, our statistical analysis does not select accentedness as a significant predictor for monophthongisation but considers variant and break type better predictors for our monophthongisation models. This is an indication that in the quotative context, *like* wound up in a small syntactic unit in which reduction is highly likely as *several* contextual factors conspire to reduce *like*. The most typical context for quotative *be like* is an unaccented *like* within a proclitic reporting construction followed by a short break and a prominent reported clause.

Thus, it is very much the context in which *like* occurs that determines its form. Wichmann (2011) has made similar proposals; specifically, regarding grammaticalisation, namely that many descriptions of reduction and elision that occur in the process of grammaticalisation are often due to the loss of prosodic prominence that then ultimately leads to segmental changes. While this is an interesting argument, it does seem to contradict the observation that discourse markers often receive tonic stress and are followed by a pause (Schiffrin 1987, cited in Aijmer 2002: 32) rather than being reduced. But Halliday & Hasan (1976: 271) observe two patterns: 'there is a general tendency in spoken English for conjunctive elements as a whole to be, phonologically, either tonic (maximally prominent) or reduced (minimally prominent) rather than anything in between'. This is in line with our findings: the evolution of *like* has led to some functions of *like* being more or less accented than others, which

[14] Even if we view *like* in the quotative as a complementiser, we would expect similar weakening. According to Kelly & Bock (1988: 393), who constructed a stress ranking for ten grammatical classes of monosyllabic words, only 14 per cent of complementisers occurred in a strong position.

sheds light on reduction processes in language change. Our data suggest that language change that involves syntactic reanalysis and increased clausal integration,[15] changes in position and large loss of structural independence, as is the case for *be like*, is likely to be subject to prosodic changes and also reduction.

This reduction-friendly context, in which *be like* occurs, trumps any frequency effects that may occur. For example, although the discourse particle is the most frequent function of *like*, it is not the most reduced function (see literature review). Nor is there an effect for speakers who are more likely to use *be like* to produce more reduced *like* tokens in this construction. Nonetheless, in our study, it is only the vowel that seems to be sensitive to the function of *like*. Why we were unable to document similar *like* function-specific effects for /l/ and /k/ is an interesting question. It is possible that such effects are too variable to rise above statistical significance. In addition, social groupings that we were unable to investigate here may very well be relevant at a very local level, similar to the communities of practice documented in Drager (2011). After all, Drager's ethnographic study resulted in data that are contextually more variable than ours and at times extremely informal. This may explain some of the differences in our findings and represents an important area of future research.

Future research may also involve different measures of predictability. As one of the reviewers has pointed out, discourse context may be an important factor by which the *be like* quotative is predictable. The quotative is associated particularly with direct reporting in narratives. This activity type as well as its lexical embedding (*be* and a quotation) make this particular *like* more predictable. This is very different from the discourse marker and the discourse particle, which occur more freely. Their occurrence is much harder to predict; consequently, as the reviewer points out, auditory cues matter more for these than for quotative *be like*.

It also remains to be investigated to what extent the 'design' of the quotative – be that *say*, *go* or *be like* – prepares the context for a particular person's voice whose speech, thoughts or actions are to be reported and the way it is characterised and stylised. Klewitz & Couper-Kuhlen (1999: 477–8) argue that 'foreshadowing' a voice lexically, prosodically or paralinguistically is not uncommon. Moreover, it is not stretching a point to assume that the choice and phonetic design of the quotative itself may contribute to this stylisation. Again, an in-depth knowledge of the community in question is key.

Finally, the precise prosodic characteristics with which the *be like* quotative occurs remain to be determined in much more detail. It is most likely a bundle of prosodic characteristics that marks reported speech and may involve not only breaks but also pitch, volume and rhythm (see Klewitz & Couper-Kuhlen 1999). This may be best explored by including in the investigation the reporting as well as the reported passage.

[15] Note that many discourse markers violate Lehmann's (1995) grammaticalisation criterion of scope. Grammaticalisation normally involves syntactic integration. However, the development of discourse markers often results in a widening of scope and a decrease in clausal integration.

## 6 Conclusion

We have shown in this study that different functions of *like* may vary in their phonetic realisation of the vocalic element of *like*. Thus, this patterning is not at all limited to English in New Zealand. At a more theoretical level, this is evidence that different lemmas with the same word form (or to use alternative terminology: different lexemes with the same stem) can vary in their phonetic realisation. Drager (2011) reaches a similar conclusion in her study, and argues that her results challenge theoretical models that assume a single phonetic representation for polysemous and homophonous words that share a word form, as these models would predict the same realisation of all functions of *like*. However, we diverge in our conclusion. We do not believe that the evidence presented provides much additional support to production models with an acoustically rich lemma level, or one that is indexed directly to acoustic information. We do not dismiss the possibility that the lemmas under consideration may be indexed to separate phonological representations. Our own data do not provide clear evidence to suggest that items are stored in the mind in such a way that functional distinctions are maintained.

Jurafsky *et al.* (2002) argue that differences in the phonetic realisation of lemmas that have the same word form are reduced or disappear once contextual predictability is considered. This certainly seems to be the case for the functions of *like* as well as /l/ and /k/, but even vowel reduction is significantly dependent not only on function, but also on contextual factors. Thus, phonetic reduction seems to be due to syntactic and prosodic constraints. We have found evidence that probability effects matter at the word level of *like*, but rarely ever at the functional level.

These findings make claims that *like* is stored in the mind in such a way that functional distinctions are maintained much less convincing. But in spite of *like* reduction being context rather than function driven, the fact that some functions of *like* occur more frequently in environments in which they are subject to reduction is interesting nonetheless. Previous research has shown that diffusion of a sound change can be affected by the relative frequency of the immediate linguistic context. If certain words occur more often in a particular environment that favours change, these words may change more quickly (Bybee 2002). Thus, if one of the *like* functions occurs consistently in a particular environment that renders *like* tokens more amenable to reduction, over time, language change may lead to a continued divergence in form. Quotative *like* may become increasingly independent of the current prosodic and syntactic constraints. Different descriptions of phonological representation may then be useful. However, for the moment, the form of *like* can be predicted from context, and any claims that involve activation of different functions of *like* without a phonological buffer strike us as premature.

*Authors' addresses*

*Department of English and American Studies*
*University of Salzburg*

*UniPark Nonntal*
*Erzabt-Klotz-Straße 1*
*5020 Salzburg*
*Austria*
*erik.schleef@sbg.ac.at*

## References

Aijmer, Karin. 2002. *English discourse particles*. Amsterdam and Philadelphia: John Benjamins.

Aitken, Adam Jack. 1984. Scottish accents and dialects. In Peter Trudgill (ed.), *Language in the British Isles*, 94–114. Cambridge: Cambridge University Press.

Andersen, Gisle. 2001. *Pragmatic markers and sociolinguistic variation*. Amsterdam and Philadelphia: John Benjamins.

Ash, Sharon. 1982. The vocalization of /l/ in Philadelphia. Unpublished PhD thesis, University of Pennsylvania.

Beaken, Michael. 1971. A study of phonological development in a primary school population of east London. Unpublished PhD thesis, University College London.

Beckman, Mary E. & Julia Hirschberg. 1993. The ToBI annotation conventions. Unpublished manuscript. Available online: http://www1.cs.columbia.edu/~jjv/pubs/tobi_convent.pdf. Last accessed: 12 August 2014.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113, 1001–24.

Berg, Thomas. 2000. The position of adjectives on the noun–verb continuum. *English Language and Linguistics* 4, 269–93.

Blyth, Carl, Sigrid Recktenwald & Jenny Wang. 1990. I'm like, 'say what?!' A new quotative in American oral narrative. *American Speech* 65, 215–27.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341–5.

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

Brugman, Hennie & Albert Russel. 2004. Annotating multi-media/multi-modal resources with ELAN. In M. Lino, M. Xavier, F. Ferreira, R. Costa & R. Silva (eds.), *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*, 2065–8. Paris: European Language Resources Association.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41, 977–90.

Buchstaller, Isabelle. 2001. An alternative view of *like*: Its grammaticalization in conversational American English and beyond. *Edinburgh Working Papers in Applied Linguistics* 11, 21–41.

Buchstaller, Isabelle. 2004. The sociolinguistic constraints on the quotative system: US English and British English compared. Unpublished PhD thesis, University of Edinburgh.

Buchstaller, Isabelle. 2006. Social stereotypes, personality traits and regional perception displaced: Attitudes towards the 'new' quotatives in the UK. *Journal of Sociolinguistics* 10, 362–81.

Buchstaller, Isabelle. 2014. *Quotatives: New trends and sociolinguistic implications*. Malden, MA, and Oxford: Wiley-Blackwell.

Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalisation: A multi-local, multivariate investigation of quotative *be like*. *Journal of Sociolinguistics* 13, 291–331.

Bush, Nathan. 1999. The predictive value of transitional probability for word-boundary palatalization in English. Unpublished Master's thesis, University of New Mexico, Albuquerque, NM.

Butters, Ronald. 1982. Editor's note to Schourup (1982). *American Speech* 57, 149.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261–90.

Bybee, Joan. 2003. Mechanisms of change in grammaticalization: The role of frequency. In Brian D. Joseph & Richard D. Janda (eds.), *Handbook of historical linguistics*, 602–23. Oxford: Blackwell.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37, 575–96.

Carter, Paul & John Local. 2007. F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association* 37, 183–99.

Cheshire, Jenny, Paul Kerswill, Sue Fox & Eivind Torgersen. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15, 151–96.

Chirrey, Deborah. 1999. Edinburgh: Descriptive material. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 223–9. London: Arnold.

Cruttenden, Alan. 2001. *Gimson's Pronunciation of English*, 6th edition. Abingdon and New York: Arnold.

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

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

D'Arcy, Alexandra. 2007. *Like* and language ideology: Disentangling fact from fiction. *American Speech* 82, 386–419.

D'Arcy, Alexandra. 2014. Discourse. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 410–22. New York: Routledge.

D'Arcy, Alexandra. 2015. Quotation and advances in understanding syntactic systems. *Annual Review of Linguistics* 1, 43–61.

Dailey-O'Cain, Jennifer. 2000. The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of Sociolinguistics* 4, 60–80.

Dion, Nathalie & Shana Poplack. 2007. Linguistic mythbusting: The role of the media in diffusing change. Presented at New Ways of Analysing Variation 36, University of Pennsylvania.

Dinkin, Aaron J. 2008. The real effect of word frequency on phonetic variation. University of Pennsylvania Working Papers in Linguistics 14.1.

Drager, Katie K. 2009. A sociophonetic ethnography of Selwyn Girls' High. Unpublished PhD Thesis, University of Canterbury.

Drager, Katie K. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39, 694–707.

Durham, Mercedes, Bill Haddican, Eytan Zweig, Daniel Ezra Johnson, Zipporah Baker, David Cockeram, Esther Danks & Louise Tyler. 2012. Constant linguistic effects in the diffusion of *be like*. *Journal of English Linguistics* 40, 316–37.

ELAN. Language archiving technology, version 3.8.0. 2009. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. http://tla.mpi.nl/tools/tla-tools/elan/. Accessed 6 November 2009.

Ferrara, Kathleen & Barbara Bell. 1995. Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *be + like*. *American Speech* 70, 265–90.

Fowler, Carol A. & Jonathan Housum. 1987. Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26, 489–504.

Fox, Sue. 2007. The demise of Cockneys? Language change in London's 'traditional' East End. Unpublished PhD thesis, University of Essex.

Haas, Florian. 2007. The development of English each other: Grammaticalization, lexicalization, or both? *English Language and Linguistics* 11, 31–50.

Haddican, William, Eytan Zweig & Daniel Ezra Johnson. 2015. Change in the syntax and semantics of *be like* quotatives. In Theresa Biberauer & George Walkden (eds.), *Syntax over time: Lexical, morphological and information-structural interactions. Proceedings of the 12th meeting of Diachronic Generative Syntax (DiDS XII)*, 54–71. Oxford: Oxford University Press.

Halliday, M. A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Halliday, M. A. K. & Christian M. I. M. Matthiessen. 2004. *An introduction to Functional Grammar*, 3rd edition. London: Arnold.

Hay, Jennifer & Harold Baayen. 2002. Parsing and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology*, 203–35. Dordrecht: Kluwer.

Heine, Bernd. 2003. Grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 575–601. Oxford: Blackwell.

Heuven, Walter J.B. van, Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. Subtlex-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67, 1176–90.

Hopper, Paul & Elizabeth C. Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15, 651–74.

Hughes, Arthur, Peter Trudgill & Dominic Watt. 2012. *English accents and dialects: An introduction to social and regional varieties of English in the British Isles*. London: Routledge.

Johnston, Paul. 1985. The rise and fall of the Morningside/Kelvinside accent. In Görlach (ed.), *Focus on: Scotland*, 37–56. Amsterdam and Philadelphia: John Benjamins.

Johnston, Paul. 2007. Scottish English and Scots. In David Britain (ed.), *Language in the British Isles*, 105–21. Cambridge: Cambridge University Press.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in laboratory phonology VII*, 1–34. New York: Mouton de Gruyter.

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 229–54. Amsterdam and Philadelphia: John Benjamins.

Kelly, Michael H. & J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14, 389–403.

Kerswill, Paul, Eivind Torgersen & Sue Fox. 2008. Reversing 'drift': Innovation and diffusion in the London diphthong system. *Language Variation and Change* 20, 451–91.

Klewitz, Gabriele & Elizabeth Couper-Kuhlen. 1999. Quote – unquote? The role of prosody in the contextualization of reported speech sequences. *Pragmatics* 9: 459–485.

Krug, Manfred. 1998. String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26, 286–320.

Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.

Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: Lincom Europa.

Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 1–18. Amsterdam and Philadelphia: John Benjamins.

Meyerhoff, Miriam & Erik Schleef. 2013. Hitting an Edinburgh target: Immigrant adolescents' acquisition of variation in Edinburgh English. In Robert Lawson (ed.), *Sociolinguistic perspectives on Scotland*, 103–28. Basingstoke: Palgrave Macmillan.

Phillips, Betty S. 2006. *Word frequency and lexical diffusion*. Basingstoke and New York: Palgrave Macmillan.

R Core Team. 2014. *R: A language and environment for statistical computing*. www.R-project.org. Vienna: R Foundation for Statistical Computing.

Rhodes, Richard A. 1996. English reduced vowels and the nature of natural processes. In Bernhard Hurch & Richard Rhodes (eds.), *Natural phonology: The state of the art*, 239–59. Mouton de Gruyter.

Romaine, Suzanne & Deborah Lange. 1991. The use of *like* as a marker of reported speech and thought: A case of grammaticalisation in process. *American Speech* 66, 227–79.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.

Schleef, Erik. 2013. Glottal replacement of /t/ in two British capitals: Effects of word frequency and morphological compositionality. *Language Variation and Change* 25, 201–23.

Schleef, Erik, Miriam Meyerhoff & Lynn Clark. 2011. Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London. *English World-Wide* 32, 206–36.

Schourup, Lawrence Clifford. 1983. Common discourse particles in English conversation. Ohio State Working Papers in Linguistics 28.

Scobbie, James M., Nigel Hewlett & Alice E. Turk. 1999. Standard English in Edinburgh and Glasgow: The Scottish vowel length rule revealed. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 230–45. London: Arnold.

Scobbie, James M. & Marianne Pouplier. 2010. The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics* 38, 240–59.

Selkirk, Elizabeth O. 1984. *Phonology and syntax*. Cambridge, MA: MIT Press.

Speitel, Hans H. 1983. A sociolinguistic investigation of Edinburgh speech. Unpublished end of grant report for the Social Science Research Council.

Sproat, Richard & Osamu Fujimora. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291–311.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14, 323–48.

Stuart-Smith, Jane. 1999. Glasgow: Accent and voice quality. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 203–22. London: Arnold.

Stuart-Smith, Jane. 2008. Scottish English: phonology. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English: The British Isles*, 48–70. Berlin and New York: Mouton de Gruyter.

Tagliamonte, Sali A. & Alexandra D'Arcy. 2004. *He's like, she's like*: The quotative system in Canadian youth. *Journal of Sociolinguistics* 8, 493–514.

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

Tagliamonte, Sali A. & Rachel Hudson. 1999. *Be like* et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3, 147–172.

Tamminga, Meredith. 2013. Sound change without frequency effects: Ramifications for phonological theory. In Robert E. Santana-LaBarge (ed.), *Proceedings of the 31st West Coast Conference on Formal Linguistics*, vol. 31, 457–65. Somerville, MA: Cascadilla Proceedings Project.

Tollfree, Laura. 1999. South East London English: Discrete versus continuous modelling of consonantal reduction. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 163–84. London: Arnold.

Turton, Danielle. 2014. An ultrasound investigation of /l/-darkening in English. University of Pennsylvania Working Papers in Linguistics 20.

Underhill, Robert. 1988. Like is like, focus. *American Speech* 63, 234–46.

Vandelanotte, Lieven. 2012. Quotative *go* and *be like*: Grammar and grammaticalization. In Ilse Buchstaller & Ingrid van Alphen (eds.), *Quotatives: Cross-linguistic and cross-disciplinary perspectives*, 173–202. Amsterdam and Philadelphia: John Benjamins.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Wichmann, Anne. 2011. Grammaticalization and prosody. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 331–41. Oxford: Oxford University Press.

Yuan, Jiahong & Mark Liberman. 2009. Investigating /l/ variation in English through forced alignment. *Interspeech* 10, 2215–18.

Yuan, Jiahong & Mark Liberman. 2011. /l/ variation in American English: A corpus approach. *Journal of Speech Sciences* 1, 35–46.

Zipf, George Kingsley. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 15, 1–95.

# Appendix

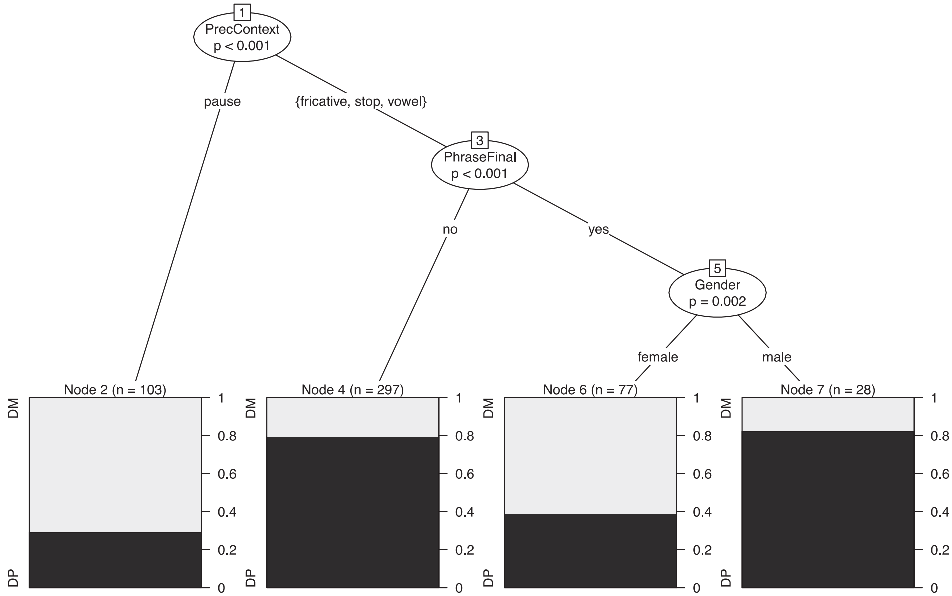


Figure A1. Edinburgh discourse marker vs discourse particle

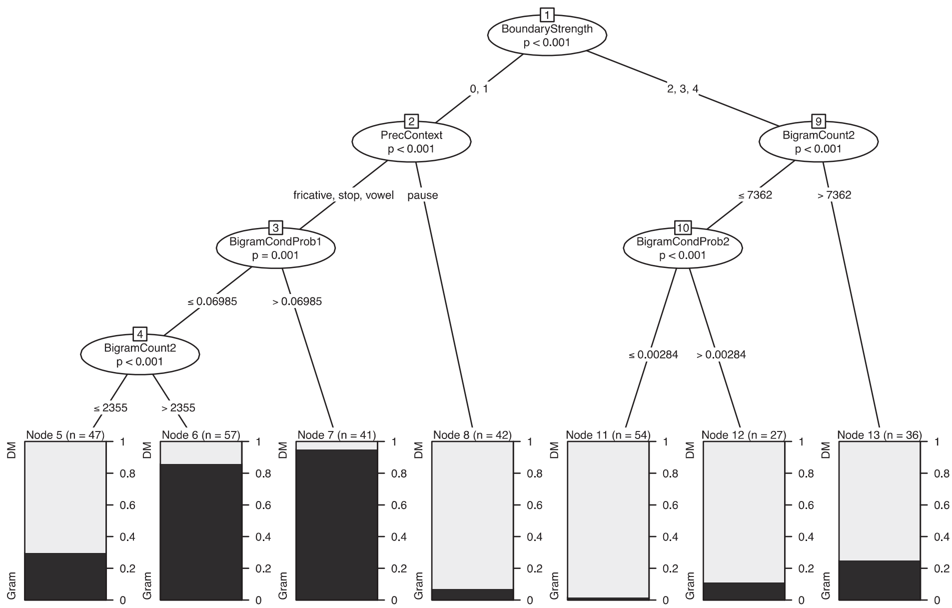


Figure A2. Edinburgh discourse marker vs grammatical

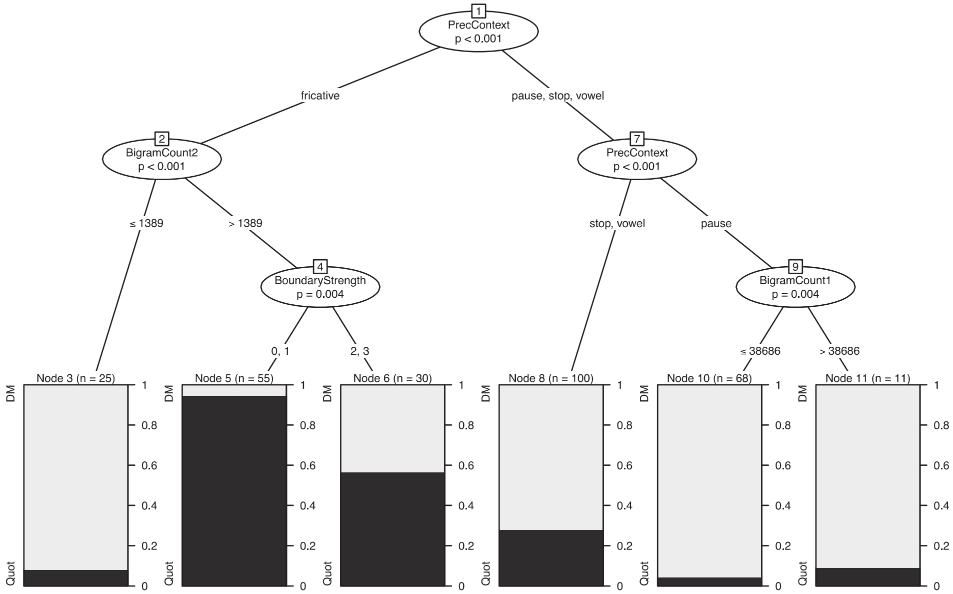


Figure A3. Edinburgh discourse marker vs quotative

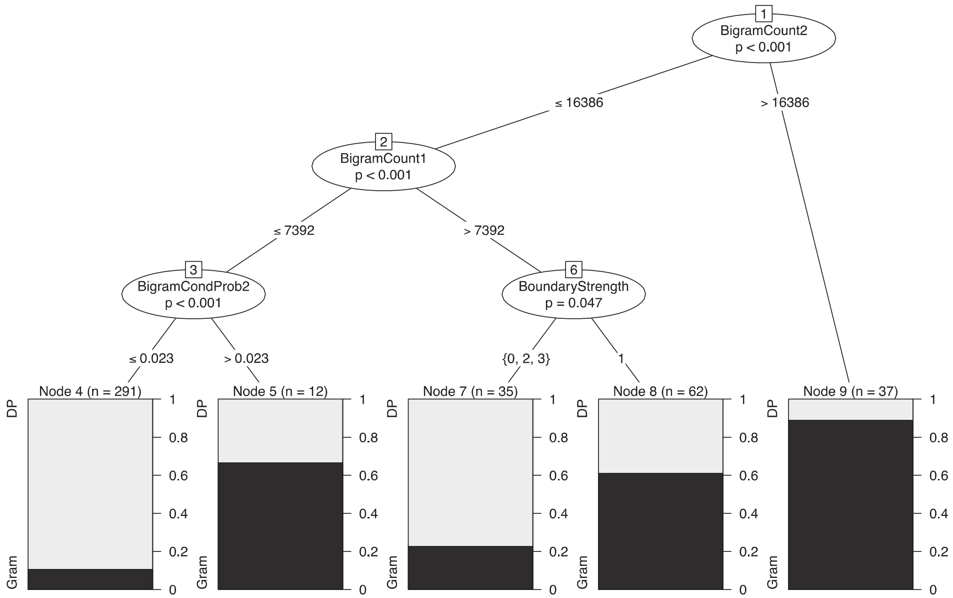


Figure A4. Edinburgh discourse particle vs grammatical

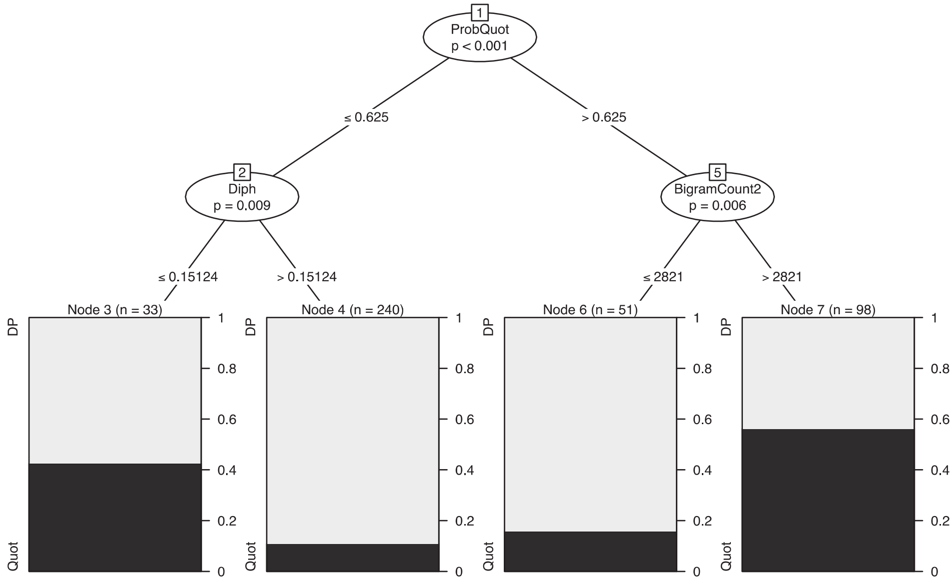


Figure A5. Edinburgh discourse particle vs quotative

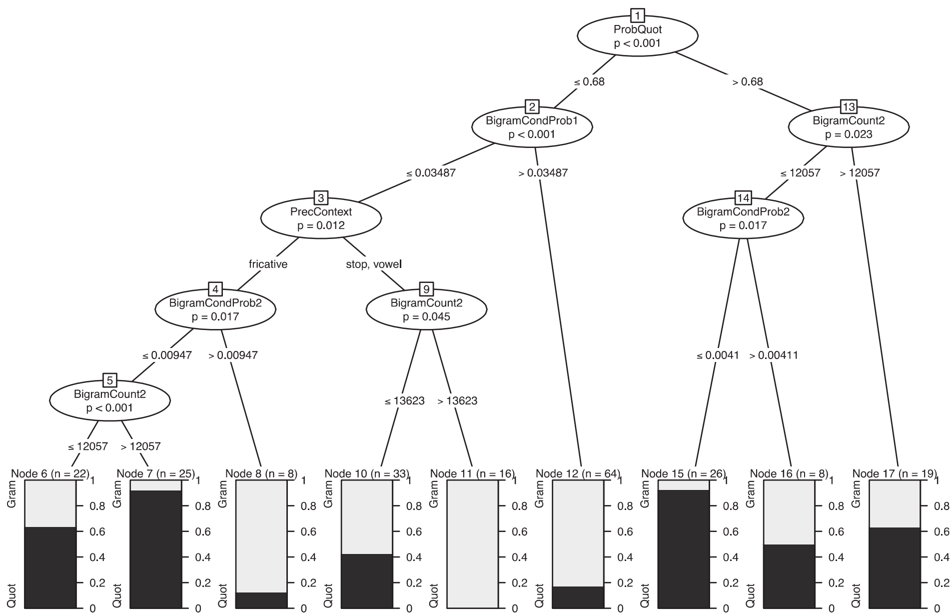


Figure A6. Edinburgh grammatical vs quotative

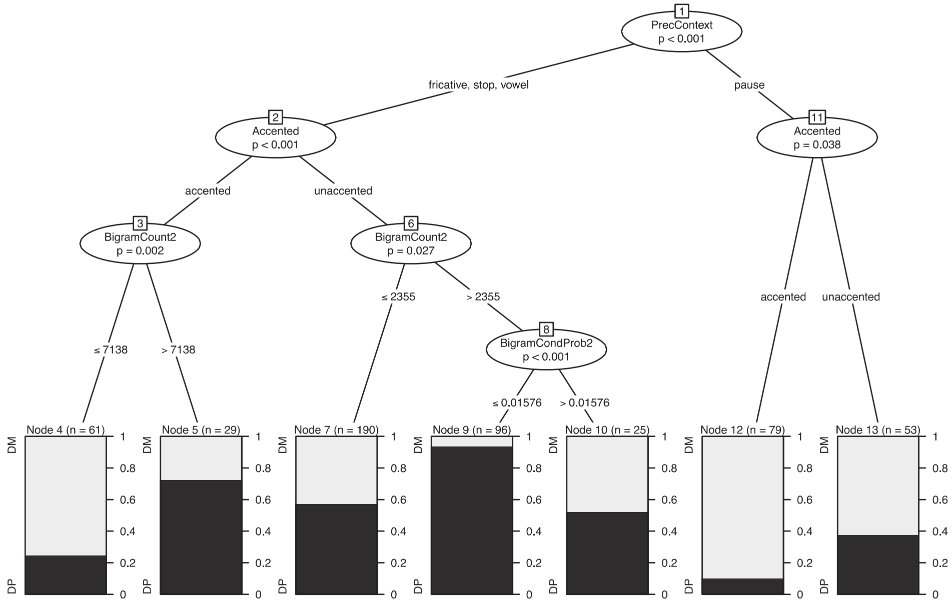


Figure A7. London discourse marker vs discourse particle

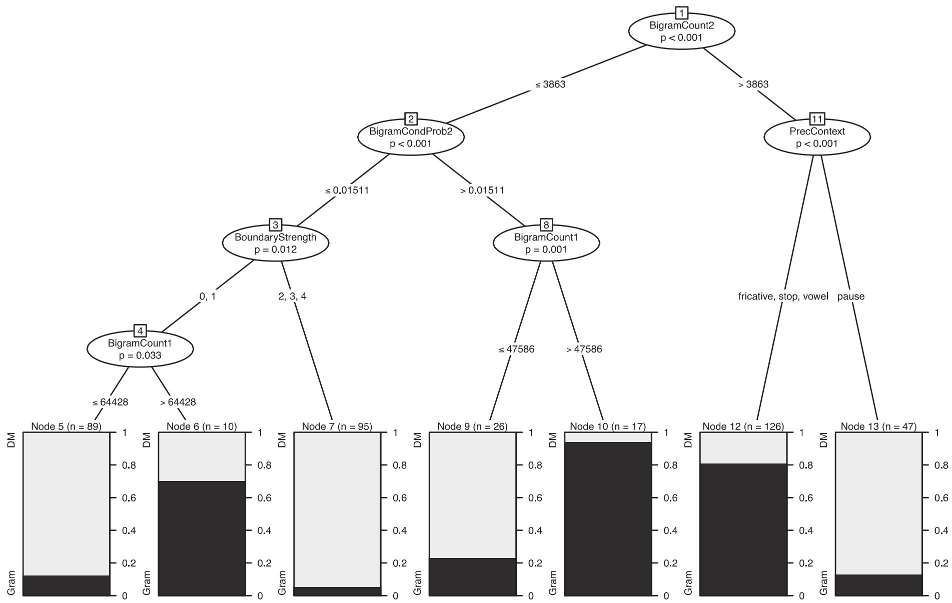


Figure A8. London discourse marker vs grammatical

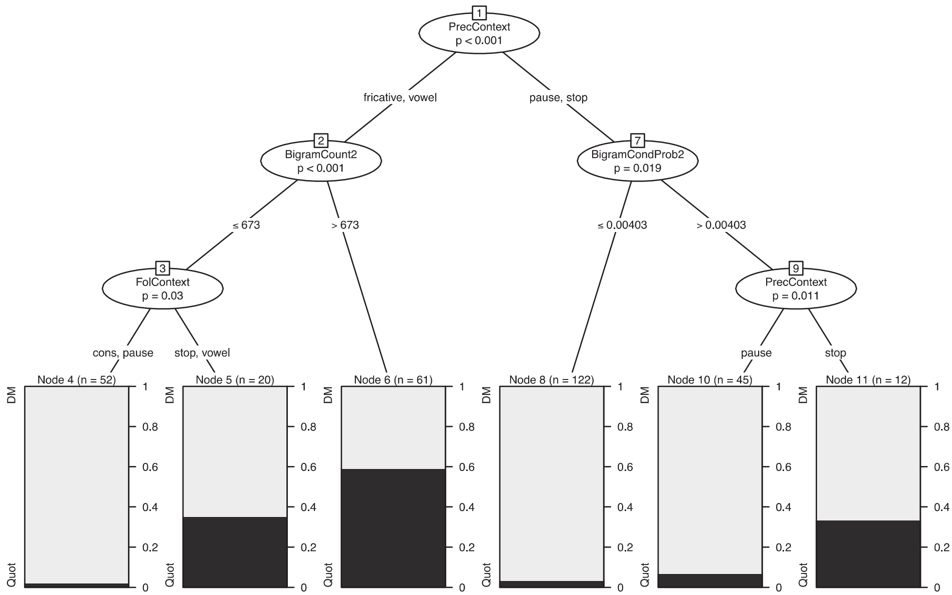


Figure A9. London discourse marker vs quotative

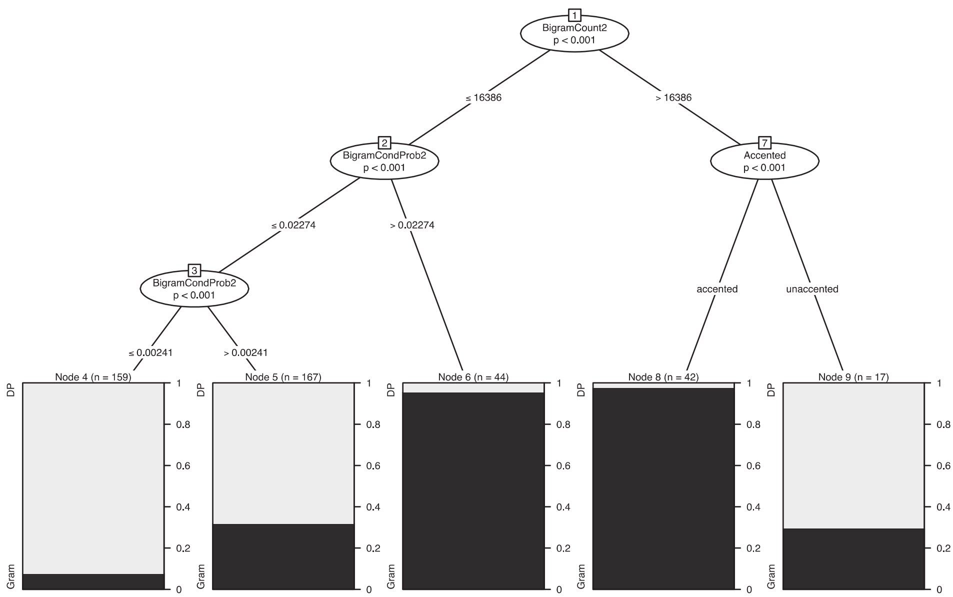


Figure A10. London discourse particle vs grammatical

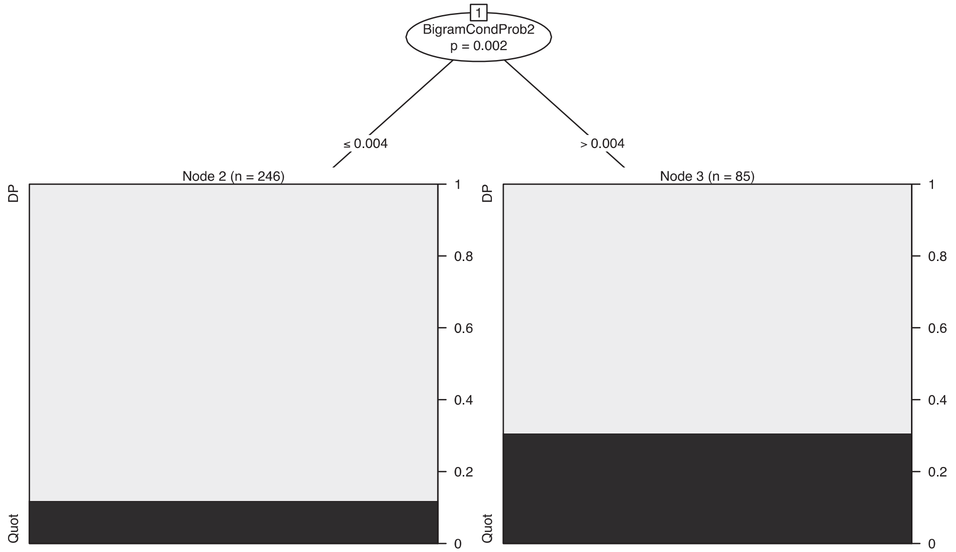


Figure A11. London discourse particle vs quotative

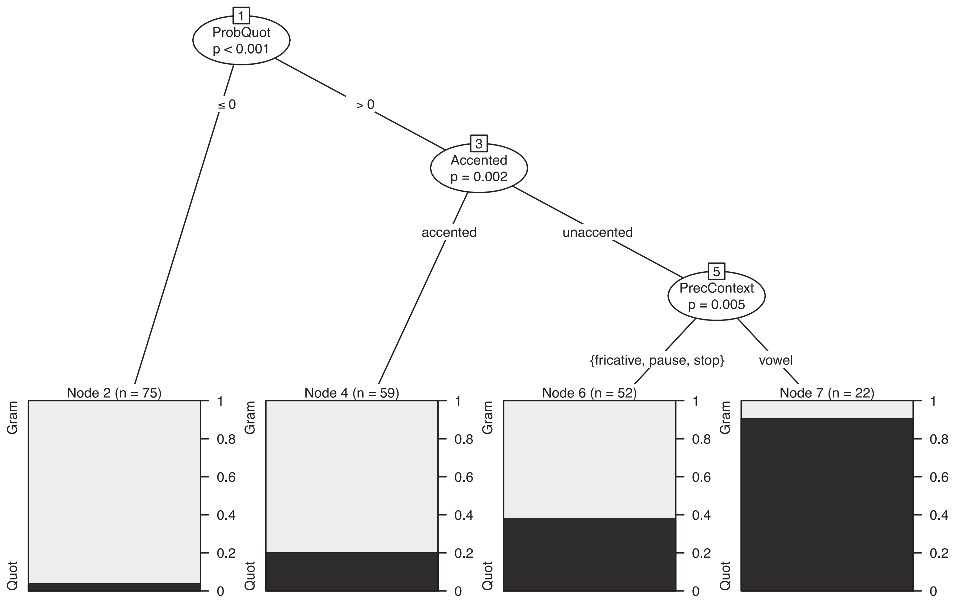


Figure A12. London grammatical vs quotative