A CORPUS-BASED ANALYSIS OF THE PRAGMATIC MARKER, YOU GET ME

EIVIND TORGERSEN is Professor of English at the Norwegian University of Science and Technology. He has worked on projects on Multicultural London English and language change in London, in particular modelling of phonological change and the use of spoken corpora in sociolinguistic research. Other research interests are in experimental phonetics and second language acquisition.

COSTAS GABRIELATOS is Senior Lecturer in English Language at Edge Hill University. His research combines the following areas: corpus linguistics methodology (topic-specific corpora, annotation techniques, metrics), grammar (conditionals, modality, tense-aspect, construction grammar, lexical grammar), language education (pedagogical lexicogrammar, learner language), and discourse studies.

SEBASTIAN HOFFMANN is Professor of English Linguistics at Trier University. His research predominantly focuses on the application of usage-based approaches to the study of language; recent research topics include: syntactic change, tag questions, the lexico-grammar of New Englishes and corpus linguistic methodology involving Internet-derived data.

Abstract

The chapter examines the use of the emerging pragmatic marker *you get me* (e.g. *I'm just gonna give her a little backhand or whatever cos she needs to learn you get me?*) in the 1.5 million word Multicultural London English Corpus (MLEC) (2008). The corpus contains

sociolinguistic interviews with London English speakers and the metadata provide information about a speaker's ethnicity, sex, and age group. The methodology combines automated and manual analysis, and draws on two related previous studies (Gabrielatos, Torgersen, Hoffmann, & Fox, 2010; Torgersen, Gabrielatos, Hoffmann, & Fox, 2011), which used the Linguistic Innovators Corpus (LIC) (2005), a 1.4 million word corpus of sociolinguistic interviews with inner- and outer-London speakers, also marked-up for ethnicity, sex and age, as well as locality. The analysis focuses on the extent of use of *you get me,* as well as its its variants and discourse functions in relation to the sociolinguistic factors outlined above. The analysis also incorporates comparisons with the use of *you get me* in LIC, in which ethnicity emerged as the strongest factor.

Introduction

For several decades now, the methodological toolkit of sociolinguists investigating phenomena of spoken language has largely been different from the approaches employed by corpus linguists. For the most part, traditional sociolinguistic analysis relies on the careful analysis of audio recordings made in the context of various types of fieldwork, e.g. the sociolinguistic interview (for an overview of typical approaches, see Labov (1984)). This analysis typically focuses on phonological or morpho-syntactic variables and requires researchers to detect relevant extracts in their data, which are then described in detail and interpreted against the background of the extensive information available for individual speakers, using highly elaborate quantitative procedures such as variable rule analysis (Sankoff, 1988). The data collected as part of such projects have rarely involved the orthographic transcription of complete recordings, but relevant extracts have been transcribed

when necessary, e.g. for conversational analysis of short passages. Examples include the Milton Keynes project (Kerswill & Williams, 2000) and the Reading-Hull-Milton Keynes dialect levelling project (Cheshire, Kerswill, & Williams, 2005).

In our paper, we build on recent developments that have led to a narrowing of the gap between traditional sociolinguistics and corpus linguistics. Thus, the studies by Tagliamonte and colleagues (1998; 2005), which analyse *was/were*-variation and relative marking in northern Britain respectively, relied on orthographic transcriptions of full interviews; Rayson, Leech & Hodges (1997) and McEnery & Xiao (2004) examined lexical use and swearing (respectively) in the spoken section of the British National Corpus in relation to sociolinguistic factors. A similar approach was taken by the two sociolinguistic projects which are at the heart of the discussion in our paper. The data comprising the two corpora used in our study were collected for the projects *Linguistic Innovators: The English of adolescents in London* (2004-2007) (henceforth LIC), which was the first large-scale sociolinguistic project set in London, and *Multicultural London English: the emergence, acquisition and diffusion of a new variety* (2007-2010) (henceforth MLEC), which built on the first project. In the former project, interview data were collected from 122 speakers, who were sampled according to age, sex, ethnicity and geographical location. In the latter, interview data were collected from 127 speakers, who were sampled according to age, sex and ethnicity. The combined transcribed datasets contain 2.8m words and contain rich levels of sociolinguistic metadata for all speakers. This is important, as detailed information about the corpus speakers has in the past sometimes been missing in spoken language corpora, such as speaker ethnicity in COLT (Stenström, Andersen, & Hasund, 2002), and indeed, has not always been considered as important when the goal was to compile a large corpus (Sinclair, 1995). All interview data for LIC and MLEC were transcribed orthographically, and analyses

3

of morpho-syntactic and discourse features were carried out. These investigations, which relied on purely manual extraction of tokens, have found effects of speaker ethnicity, friendship network and geographical location, pointing to linguistic innovation taking place in inner London.

We would like to show that the methodological gap between sociolinguistics and corpus linguistics can be further closed. In particular, this paper discusses how insights from sociolinguistics, specifically the inclusion of speaker ethnicity and sex (Milroy, 1987) and their interaction, can improve the utility of the corpus in explaining linguistic processes of variation and change. Specifically, we analyse the sociolinguistic distribution and functions of a discourse feature, the pragmatic marker *you get me* as used by young speakers in inner London. We document that speakers' ethnicity needs to be discussed in relation to language change and innovation in large urban centres and must also be included in the sampling of corpus speakers.

The Linguistic Innovators project: overview and findings

The research project *Linguistic Innovators: The English of Adolescents in London* (Kerswill, Cheshire, Fox, & Torgersen, 2007) was set up to investigate whether London was the source of language changes in Britain, as has been suggested by e.g. Wells (1982). The main research questions were:

- Where do linguistic innovations take place?
- How do they spread?
- Who are the linguistic innovators?

Given these questions, it is clear that answers should be sought among external linguistic factors such as geographical location, gender and age. To answer these questions, two

research sites were chosen. One was in inner London, Hackney, the other in outer London, further to the east, Havering, as shown in Figure 1.

Figure 1 Map of localities in the Linguistic Innovators project.

The sample of speakers from the inner city area consists of adolescent speakers aged 16-19. Half of them have a 'white London' background, in that their families have lived in the area for at least three generations. This group of speakers was termed 'Anglo.' The remaining half are the children or grandchildren of immigrants and were termed 'non-Anglo'. Havering has a very different ethnic mix and is one of London's least diverse boroughs, but it has recently seen its ethnic-minority population doubled. The white British population is now down to 83 per cent (Office for National Statistics, 2016). The sample for Havering therefore consisted of adolescent speakers almost all of white British background. In addition, working-class Anglo adults aged 65-80 in Hackney and Havering were recorded, to act as a reference point for comparison with the speech of the adolescents.

The main finding was indeed that social variables play a large role in language change and innovation in London English. Significant effects for phonological and grammatical change were living in an inner city location (Hackney), having a Non-Anglo ethnicity, being in a dense multi-cultural friendship network, and being male. The non-Anglo male speakers from Hackney were in the lead in segmental, suprasegmental and morpho-syntactic changes, such reversing the process of diphthong shifting (Kerswill, Torgersen, & Fox, 2008), having a more syllable-timed speech rhythm (Torgersen & Szakay, 2012) and using the indefinite article form *a* before vowel-initial words (Gabrielatos, Torgersen, Hoffmann, & Fox, 2010). Hackney non-Anglo speakers tended to favour levelling to *was* in positive polarity and levelling to *wasn't* in negative polarity contexts (Cheshire & Fox, 2009). Speakers in dense multi-cultural friendship networks had a specific use of relative pronouns, namely of *who* as topic marker in restrictive relative clauses, such as 'my medium brother who moved to Antigua' (Cheshire, Adger, & Fox, 2013, p. 72) and Anglo speakers who were not in such friendship networks has less use of the pragmatic marker *you get me* than Anglo speakers who were in such networks (Torgersen, Gabrielatos, Hoffmann, & Fox, 2011).

While the project documented the role of ethnicity and friendship network in linguistic innovation and Multicultural London English as a variety spoken by young speakers in Hackney of different ethnic backgrounds, the possible origin of the variety, how it is acquired and whether it is spoken in other parts of London were not documented.

The Multicultural London English project: overview and findings

The main objective for the project *Multicultural London English: the emergence, acquisition and diffusion of a new variety* (Kerswill, Cheshire, Fox, & Torgersen, 2011) was to examine

the acquisition of Multicultural London English (MLE) by children younger than the adolescents interviewed in the *Linguistic Innovators* study. The main research questions were:

- What are the origins of MLE?

- When and how is MLE acquired by younger children?

- Is MLE spoken outside of Hackney?

Data came from different age groups: four-, eight-, twelve- and sixteen-year olds, as well as speakers in their mid 20s. The latter group was included to examine whether MLE features are maintained into adulthood. In addition, the parents of the youngest children were recorded to examine linguistic transmission, which is the passing-on of linguistic features from one generation to the next. Speakers were divided into two broad ethnic groups as in the *Linguistic Innovators* project: Anglos and non-Anglos. Data were collected in areas to the west and north of the inner-London location in the *Linguistic Innovators* project, the boroughs of Haringey and Islington, in addition to Hackney itself, as shown in Figure 2. All these boroughs are considered inner-city locations with an ethnically diverse population.

Figure 2 Map of localities in the Multicultural London English project.

Studies of both phonological and grammatical features show that MLE is spoken outside of Hackney (Cheshire, Kerswill, Fox, & Torgersen, 2011). The studies examined the linguistic features already documented in the *Linguistic Innovators* project, but also revealed features which had not previously been investigated. The findings showed that there is no big difference between Anglo and non-Anglo speakers in the use of fully established features, such as H-reinstatement (speakers have less H-dropping), but a difference was found between Anglos and non-Anglos in the use of an emerging or innovative feature, K-backing, which is backing of /k/ in front of non-high vowels in words like *cousin* and *car*. The non-Anglo speakers – in particular non-Anglo teenagers – used K-backing significantly more than the Anglo speakers (Fox & Torgersen, forthcoming). An innovative pronoun *man* is mainly used by male speakers of non-Anglo background (Cheshire, 2013). Young children acquire MLE features in peer groups, i.e. their linguistic features do not reflect the use of their parents, and there is a higher frequency of MLE features with increased age. The full set of MLE features

is achieved in teenage years. Ethnicity is again documented as factor in the use of innovative linguistic features.

You get me as a pragmatic marker

This study examines the extent of use of the pragmatic marker _you get me_, its discourse functions and sociolinguistic distribution in terms of ethnicity and sex. Pragmatic markers are discourse elements typical of communication; they deal specifically with speaker attitudes and evaluation of the message content (Andersen, 2001, p. 22). They are closely linked to how a speaker wishes the interlocutors to interpret the meaning in the context in which it was uttered, as well as functioning as "contextual coordinates" (Schiffrin, 1987, p. 326) that "bracket units of talk" (Schiffrin, 1987, p. 35). There is therefore a clear focus on interaction. Torgersen et al. (2011, p. 96) argue that pragmatic markers "indicate speaker-sanctioned places in the discourse where the interlocutor can comment". Our examination of _you get me_ in LIC and MLEC shows that speakers indeed offer comments in some cases, but not in others. Instead, the speaker may expand on something that has just been said, offer clarification, or say _you get me_ as a comment on something that has been said by another speaker, irrespective of whether this comment comes in a separate turn or in a backchannel. Extracts (1) to (4) from LIC and MLEC display typical uses of _you get me_ in our data:

(1)   Dave: yeah and that they call me a mummy's boy . I don't care . it's my mum _you get me_ . [Sue: mm] call me what you want .. I'm the one that's still at home . all the luxuries and they're out there . no money yeah each week . scraping through .

(2)   Ferda: <tsk> but she looked like twenty . cos if she was even though she was still

thirteen even though I knew she was I would still go for her .. yeah she might be

thirteen but she's got the mouth . *you get me*?

Chelsea: true

Lucinda: true ..

(3)   David: I don't care bruv .. *you get me*? that's how cowardly you are you gonna stab me

over a phone

(4)   Omar: I see where you're coming from and I see [David: *you get me*] I see where

you're coming from but

In examples (1) and (2) *you get me* functions as an agreement marker where the interlocutors

offer comments, i.e. it receives a response in a backchannel (1) or a separate turn (2), while in

example (3) the speaker offers clarification without receiving a response from the interlocutor.

The distinction between uses that receive a response from an interlocutor and uses that do not

is important in the development of a linguistic item/expression towards becoming a pragmatic

marker: hearers do not (necessarily) comment when an expression has become a pragmatic

marker (Traugott, 2010; Fitzmaurice, 2004). Example (4), finally, is interesting because it is a

comment on something another speaker has said. This is an example of *you get me* losing its

internal structure involving loss of semantic content which is part of the grammaticalisation

process expressions undergo as part of becoming pragmatic markers (Lehmann, 1985). It

resembles the functional classification 'follow-up' as discussed for *innit* in COLT (Stenström

et al., 2002, p. 140).

It has been shown that pragmatic markers can be borrowed from other languages (Foolen, 2012). An alternative account is that linguistic elements from different languages and different varieties of English go into a feature pool as result of dialect and language contact (Mufwene, 2001) from where speakers can select variants based on their friendship network, ethnic background, and age (Cheshire et al., 2011). Instances of *you get me* in London English may be picked from a feature pool, and may then be further developed in MLE. It is unclear what the origin of *you get me* is, but Tormei (2015) has documented its use in adolescent speak in a Rastafarian community in Ethiopia. This is a community with immigration from Jamaica, whose dialect shows Jamaican and American English influence, in particular via rap music. His extracts of transcribed data include uses of the pragmatic markers *you see me*, *you know what I'm saying* and *you get me*, but the data is not quantified and the use of pragmatic markers is not commented on. The instances of *you get me* are in fact classified as American English (but it is not specified what variety of American English it is), e.g. "I have to try different ways of [Standard English], you get me [American English], to make it in life [Standard English], you get me [American English]" (Tormei, 2015, p. 138). However, we believe it is unclear whether *you get me* is an American 'rap import' or a linguistic innovation in MLE. What we do know, however, is that *you get me* is currently developing/emerging as a pragmatic marker in teenage MLE speakers in Hackney (Torgersen et al., 2011), but not whether it is used by speakers outside of Hackney, at what age it is acquired, and how it used. Adolescents are innovative in their use of discourse features such the *BE like* quotative (Tagliamonte & D'Arcy, 2004) and pragmatic markers such as *innit* and *like* (Andersen, 2001) and we should therefore turn to these speakers to examine the use of *you get me* in more detail.

Methodology

The study compares the use of *you get me* in two corpora, MLEC and LIC. However, in their full form, the two corpora are not comparable, as they contain different age groups and localities. To be able to compare the use of *you get me* in the datasets, two sub-corpora were created: MLEC2 consists of only the teenage speakers in MLEC, whereas LIC3 consists of only the Hackney teenage speakers in LIC (the LIC2 sub-corpus was created in Torgersen et al. (2011, pp. 98-100)) for comparisons with a sub-corpus of COLT (Stenström et al., 2002). The sociolinguistic distribution of speakers in MLEC2 and LIC3 is shown in Table 1.

Table 1 The sub-corpora MLEC2 and LIC3 used for the analysis of *you get me.*

	MLEC2	LIC3
No. of words	194,236	457,812
No. of speakers	25	51
Data collection period	2008	2005
Data collection method	Sociolinguistic interviews	Sociolinguistic interviews
Age	16-19 (average 17)	16–18 (average 17)
Sex	female; male	female; male
Ethnicity	Anglo; non-Anglo (but different ethnicities from LIC)	Anglo; non-Anglo
Residence	North and West of Inner London (Hackney, Haringey, Islington)	Inner London (Hackney)
Social class	Working class	Working class

Concordances of candidate instances of *you get me*, together with metatextual data about the speakers, were automatically extracted from the corpus by means of a Perl script. The advantage of using Perl over standard concordancing tools such as AntConc (Anthony, 2014) or Wordsmith (Scott, 2016) is that the script allowed us to identify candidate instances of *you get me* according to sociolinguistic factors. The instances were then examined to establish whether the candidates retrieved automatically were indeed used as pragmatic markers. This mostly involved examining candidate instances within a co-text of 250 characters on either side of each instance. However, in some cases it was essential to examine a wider co-text, or consult the original recordings. Quick navigation to the relevant place in a recording was facilitated by the original transcription files being time-aligned with timestamps at the beginning of a new turn, as well as at longer pauses inside turns.

The corpus instances of tokens were tabulated together with codes denoting sex and ethnicity. The quantitative analysis employed two complementary metrics: normalised frequency (number of occurrences per million words) and spread. Spread is the proportion of corpus speakers who used *you get me*, even if only once (Gabrielatos et al., 2010, pp. 306-309). The metric of spread is useful, because the high frequency of *you get me* in the corpus may not be the result of its frequent use by all/most corpus speakers, but the result of the very high frequency of use by a small minority of corpus speakers. Although each of the metrics is useful in itself, their combination affords a more nuanced picture. The interaction of frequency and spread can be visualised by plotting the values of the two metrics on a graph (for details, see Torgersen et al., 2011, pp. 100-102). The analysis also involved comparisons of the frequency and spread values of the attributes within each of the binary sociolinguistic factors (e.g. between the attributes 'female' and 'male' within the factor of age). In such comparisons, the size of the difference was also tested for statistical significance using the

log-likelihood statistic – with *p*=0.05 (G2=3.84) as the threshold for statistical significance, and *p*≤0.01 (G2≥6.63) treated as showing high statistical significance. Calculations for effect-size (frequency differences) and their statistical significance were carried out using Paul Rayson's spreadsheet (http://ucrel.lancs.ac.uk/people/paul/SigEff.xlsx).

Results and Discussion

This section will report on the analysis of the use of *you get me* in MLEC2 and LIC3. The first part of the analysis will briefly examine the use of the pragmatic marker (PM) in MLEC (i.e. the full dataset), in order to establish its distribution regarding age. The second part will then examine the PM in two respects: a) the variants (i.e. different forms) of the multi-word PM, and b) the clusters that *you get me* forms with other PMs. The third part will examine the frequency and spread of all forms of 'you get me' collectively in the two corpora. The three above parts will report on the distribution of the PM with regard to the sex and ethnicity of speakers. The fourth part will examine, a) the position of the PM in relation to the utterance it refers to and its place in the speaker's turn, and b) whether the PM received a response, in relation to its position in the utterance and turn. This can provide tentative indications regarding the pragmatic functions of *you get me*.

Distribution of use in MLEC according to age

Table 2 shows that the use of *you get me* is clearly a feature of teenage language. It has no corpus uses among the 4- and 8-year olds, and it is rarely employed by the group of 12-year olds. MLEC teenagers have almost three times the frequency and almost 50% higher spread

compared to the young adults, and almost thirty times the frequency and five times the spread of the twelve-year olds.

Table 2 Use of *you get me* in MLEC according to speakers' age.

Age group	Freq. raw	No. of words	*Freq. pmw*	Users	Speakers	*Spread*
4 years old	0	42,299	*0*	0	18	*0*
8 years old	0	102,972	*0*	0	20	*0*
12 years old	3	128,723	*23.3*	3	27	*11.1*
Teenager	124	194,236	*638.4*	14	25	*56.0*
Young adult	14	63,637	*220.0*	3	8	*37.5*

Variants

The manual examination of candidate instances of *you get me* helped identify a number of variants of the PM (see Table 3). In both corpora, the vast majority of tokens are 'you get me', and the proportions of variants are almost identical (MLEC2: 87.1%, LIC3: 87.9%). Looking at the results from another angle, more than one in ten tokens in either corpus is a variant: MLEC2: 12.9% (16 tokens), LIC3: 12.1% (17 tokens). The syntactically full form ('do you get me') represents a minority use in both corpora (MLEC2: 9.7%, LIC3: 6.4%).

Table 3 Variants of *you get me* in MLEC2 and LIC3.

Variant	MLEC2 (N=124)	*MLEC2 %*	LIC3 (N=141)	*LIC3 %*
you get me	108	*87.1*	124	*87.9*
do you get me	12	*9.7*	9	*6.4*
do you getting me	0	*0*	1	*0.7*

if you get me	0	*0*	1	*0.7*
get me	4	*3.2*	6	*4.3*

The analysis regarding users' ethnicity and sex reveals some interesting patterns (see Tables 4 and 5), although the raw frequencies involved are admittedly very low. It is clear that, in both corpora, the vast majority of variants is used by non-Anglo speakers (MLEC2: 93.8%, LIC3: 82.3%). However, the distribution of use in terms of sex is very different in the two corpora. In LIC3, males use the overwhelming majority of variants (70.6%), whereas in MLEC2 the use is balanced (50%). Also, the proportion of variants used by female speakers is 70.1% higher in MLEC2 than in LIC3.

Table 4 Distribution of variants in MLEC2 and LIC3 according to users' ethnicity.

	MLEC2 (N=16)	LIC3 (N=17)	MLEC2 %	LIC3 %
Anglo	1	3	*6.2*	*7.7*
Non-Anglo	15	14	*93.8*	*82.3*

Table 5 Distribution of variants in MLEC2 and LIC3 according to users' sex.

	MLEC2 (N=16)	LIC3 (N=17)	MLEC2 %	LIC3 %
Female	8	5	*50.0*	*29.4*
Male	8	12	*50.0*	*70.6*

A more nuanced picture emerges when we examine the use of variants in terms of combinations of factors (Table 6). Among users of variants, non-Anglo males show the highest proportion of use in both corpora, and have very similar proportions: at least half of

the users of variants are non-Anglo males (MLEC2=50%, LIC3=52.9%), with non-Anglo females having the second highest proportion of variants among users (MLEC2=43.8%, LIC3=29.4%). The above observations suggest that ethnicity influences the use of variants much more than sex, which has already been documented in the use of innovative phonological features (Kerswill et al., 2008) and grammatical and other discourse features (Cheshire et al., 2011) in MLE.

Table 6 Use of variants in MLEC2 and LIC3: combinations of sociolinguistic factors.

	MLEC2 (N=16)	LIC3 (N=17)	MLEC2 %	LIC3 %
Anglo Female	1	0	*6.2*	*0.0*
Anglo Male	0	3	*0*	*17.6*
Non-Anglo Female	7	5	*43.8*	*29.4*
Non-Anglo Male	8	9	*50.0*	*52.9*

In the rest of the paper, we will examine all the forms of the PM together, collectively referring to them as YGM.

YGM co-occurring with other PMs

Both corpora have a very small number of instances in which YGM is clustered with another pragmatic marker, i.e. it occurs immediately adjacent to another pragmatic marker. Such clustering is well known in discourse (Maschler, 1994). LIC3 has twice the proportion of clusters compared to MLEC2 (5% and 2.4% respectively), but the proportion of speakers who use clusters (i.e. spread) is slightly higher in MLEC2 (12%) than in LIC3 (9.8%). Also, in both corpora, all users of clusters are non-Anglo, with the large majority being male (66.7%

in MLEC2 and 60% in LIC3). However, there are differences in two respects. First, in MLEC2, YGM clusters with the PMs *like* (twice) and *though* (once), whereas in LIC3 it clusters with *like* (four times) and *yeah* (three times). Second, there are differences in the relative position of the PMs in the cluster. In MLEC2, YGM is always is first position (as in example (5) below), whereas in LIC3, it is either in first position (three times) or in second (four times), as in examples (6) and (7) below:

(5)   Dexter: no she's younger ... one of my sisters lives in south London ... ***you get me though***

       Aimee: uh-uh

(6)   Tina: yeah you always see the nice ones and they've got horrible girlfriends/ it's like "what you're doing with her man?"/

       Ahmed: /**yeah** *you get me* . they've got butters / girlfriends and then he's there . hot

(7)   Dom: and then because they see me . quiet like you said I'm a big guy innit and they say "oh this boy come to think that he's quite tough and then let me prove him wrong" . **like** *you get me* so they come and say something . they see me alone but if they if I do . if they do get to know me like . all about where I am from and how I am and everything I think they wouldn't say nothing they would just walk on

       Sue: mhm [Dom: yeah] so who do you hang around with what sort of people do you

Neither corpus contains enough instances for a more detailed analysis; however, the clustering of YGM, and the patterns found in the corpora seem worth investigating using a larger dataset as it will reveal more about the pragmatic functions of *you get me* when used together with another pragmatic marker.

All variants: analysis of frequency and spread

MLEC2 speakers clearly show higher use of YGM compared to LIC3 (Table 7 and Figure 3): MLEC2 has more than twice the frequency (G2=33.79) and 78.3% higher spread (but G2=2.44). However, we should look into the use in terms of sociolinguistic factors to establish the extent to which the differences are due to the sex (Table 8 and Figure 4) or ethnicity of speakers (Table 9 and Figure 5).

Table 7 Frequency and spread of YGM in MLEC2 and LIC3.

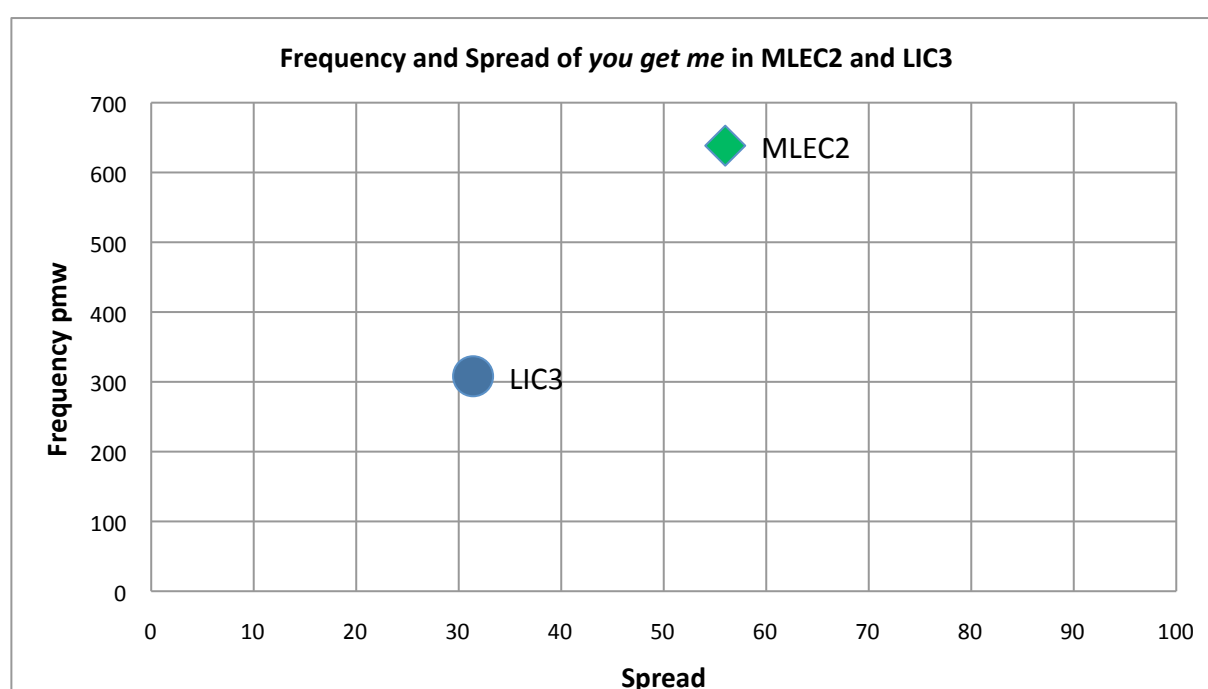|	Freq. raw	No. of words	*Freq. pmw*	Users	Speakers	*Spread*
MLEC2	124	194,236	*638.4*	14	25	*56.0*
LIC3	141	457,812	*308.0*	16	51	*31.4*

Figure 3. Frequency and spread of YGM in the two corpora.

Table 8 Frequency and spread of YGM in MLEC2 and LIC3 by sex.

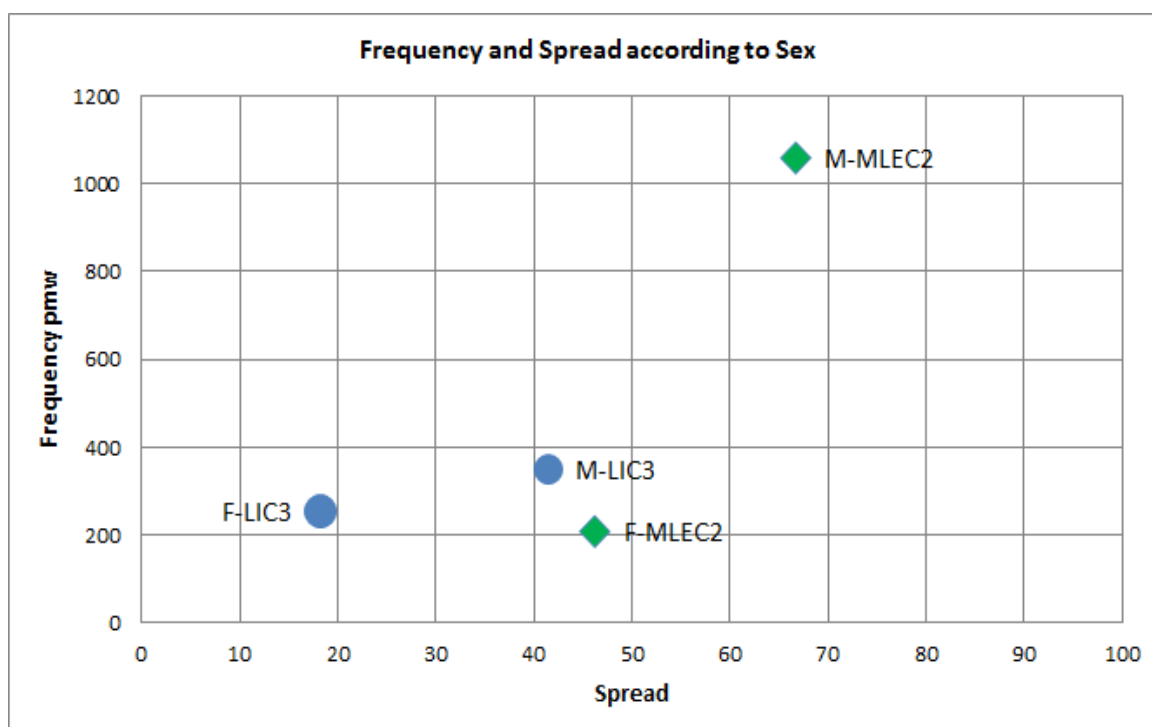|		Freq. raw	No. of words	*Freq. pmw*	Users	Speakers	*Spread*
MLEC2	**Female**	20	96,022	*208.3*	6	13	*46.2*
	Male	104	98,214	*1058.9*	8	12	*66.7*
LIC3	**Female**	50	196,776	*254.1*	4	22	*18.2*
	Male	91	261,036	*348.6*	12	29	*41.4*

Figure 4 Frequency and spread according to sex.

Examining the use according to sex within each corpus shows that, in both corpora, male speakers have higher frequency and spread than female speakers. In MLEC2, males have almost four times the frequency of females (408.4%, G2=60.45), and 44.4% higher spread (but G2=0.47). In LIC3, the frequency difference is much smaller (37.2%), and not

statistically significant (G2=3.32). However, in LIC3 the spread difference is three times higher than in MLEC2: male speakers have more than twice the spread (but G2=2.28).

Comparisons between the two corpora show that the largest difference by far is in the frequency of use of male speakers. Male speakers have three times higher frequency (203%, G2=58.41) and about 60% higher spread in MLEC2 than in LIC3 (but G2=1.05). The frequency of use by female speakers is fairly similar in the two corpora – the difference is small (22%) and not statistically significant (G2=0.58). However, female MLEC2 speakers have more than twice the spread of female LIC3 speakers (but G2=2.14). More importantly, female MLEC2 speakers have similar spread with male LIC3 speakers (in fact, slightly higher).

If any conclusions regarding development can be made by comparing MLEC2 and LIC3, the above findings suggest the following. The proportion of female users of YGM seems to be increasing, but the overall frequency appears to remain at the same level (the small decrease is not statistically significant). In males, both frequency and spread seem to be increasing – particularly the former. Simply put, more female speakers use YGM, but, collectively, less frequently, whereas moderately more males use the PM, and, collectively, much more frequently. Overall, YGM remains clearly a characteristic of male rather than female speech.

Regarding ethnicity, in both corpora, non-Anglo speakers have clearly higher frequency and spread than Anglo speakers (Table 9 and Figure 5). Frequency differences according to ethnicity are particularly pronounced: non-Anglo speakers have more than twelve times the frequency of Anglo speakers in MLEC2, and four times in LIC3 – with both differences having very high statistical significance (G2=56.25 and G2=37.64, respectively). Spread differences are smaller, but still clear, particularly in LIC3: non-Anglo speakers have

50% higher spread in MLE2, and about three times higher in LIC3 – but the differences are not statistically significant (G2=0.31 and G2=3.30 respectively).

Table 9 Frequency of YGM in MLEC2 and LIC3 by ethnicity.

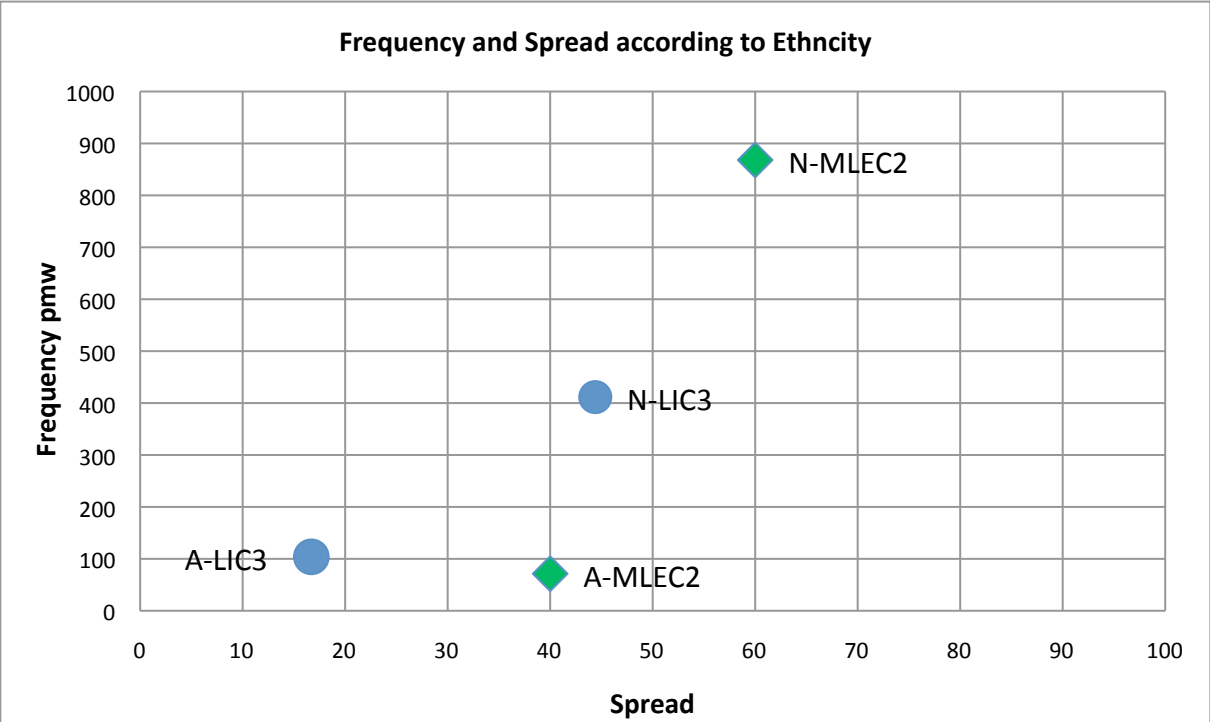|		Freq. raw	No. of words	Freq. pmw	Users	Speakers	Spread
MLEC2	**Anglo**	4	56,010	71.4	2	5	40.0
	Non-Anglo	120	138,226	868.1	12	20	60.0
LIC3	**Anglo**	16	154,019	103.9	4	24	16.7
	Non-Anglo	125	303,793	411.5	12	27	44.4

Figure 5 Frequency and spread according to ethnicity.

Comparing the two corpora, the following patterns can be identified. Anglo speakers have a fairly similar frequency in MLEC2 and LIC3. However, the spread of Anglo speakers is more

than twice as high in MLEC2, but the difference is not statistically significant (G2=0.91). MLEC2 non-Anglo speakers have twice the frequency of LIC3 non-Anglos (G=33.20), and about a third higher spread (but G2=0.54). To the extent that comparisons between MLEC2 and LIC3 can show development in the use of the PM, the above observations might be seen as indications of the following. A larger proportion of Anglo speakers seems to be adopting the PM, although they collectively still use it infrequently compared to non-Anglo speakers. The frequency among non-Anglo speakers seems to be increasing rapidly; however, spread seems to be increasing much more slowly.

Again, a more nuanced picture emerges when we examine combinations of factors (Table 10 and Figure 6). The most striking result is the extremely high frequency and spread of MLEC2 non-Anglo males, 77.8% of whom use YGM: they have almost four times higher frequency and 50% higher spread than LIC3 non-Anglo males. Regarding ethnicity, we observe the following clear pattern in both corpora: non-Anglo speakers have overwhelmingly higher frequency and spread than Anglo speakers of the same sex (e.g. non-Anglo males have higher frequency and spread than Anglo males). The only exception is MLEC2 Anglo females, who have similar spread with non-Anglo females. Regarding sex, male speakers have higher frequency and spread than female speakers of the same ethnicity. The two exceptions are MLEC2 Anglo females, who have 50% higher spread than Anglo males, and LIC3 non-Anglo females, who have almost the same frequency with non-Anglo males.

Table 10 Combinations of sociolinguistic factors in MLEC2 and LIC3.

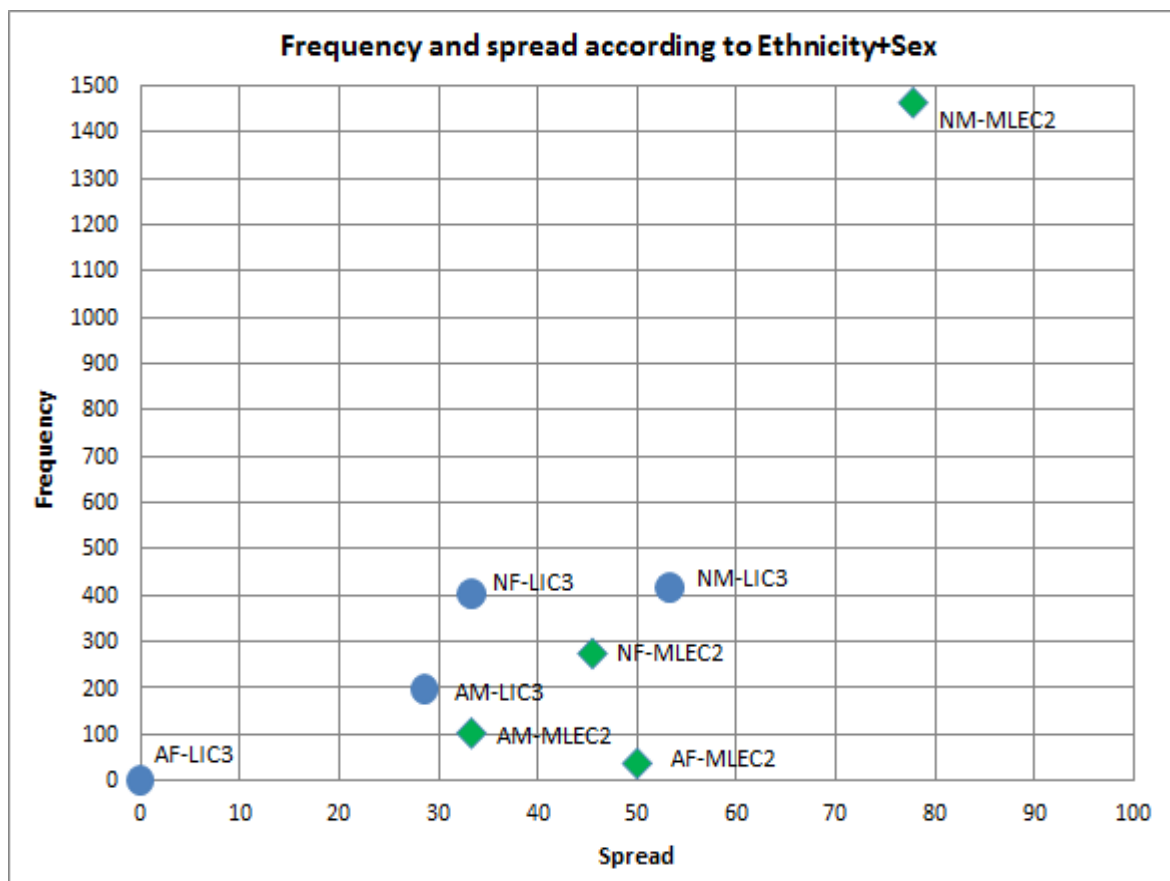|		Freq. raw	No. of words	Freq. pmw	Users	Speakers	Spread
MLEC2	Anglo Female	1	26,827	37.3	1	2	50.0
	Anglo Male	3	29,183	102.8	1	3	33.3
	Non-Anglo Female	19	69,195	274.6	5	11	45.5
	Non-Anglo Male	101	69,031	1463.1	7	9	77.8
LIC3	Anglo Female	0	72,759	0	0	10	0.0
	Anglo Male	16	81,260	196.9	4	14	28.6
	Non-Anglo Female	50	124,017	403.2	4	12	33.3
	Non-Anglo Male	75	179,776	417.2	8	15	53.3

Figure 6 Frequency and spread in MLEC2 and LIC3: combinations of sociolinguistic factors.

The main observation here is that YGM is predominantly favoured by non-Anglo males. In either corpus, they have the highest normalised frequency and spread – particularly in MLEC2. However, the use of YGM would seem to be more a characteristic of ethnicity than sex, particularly in MLEC2. Comparisons of frequency and spread within factors (female compared to male, Anglo compared to non-Anglo) have shown that, in both corpora, differences are clearly larger in terms of ethnicity than sex. Finally, comparisons of spread between the two corpora suggest that YGM is adopted by female and Anglo speakers at a higher rate compared to male and non-Anglo speakers, respectively. In other words, it seems to be increasingly adopted by speakers other than the "linguistic innovators" (non-Anglo speakers in general, and non-Anglo males in particular) (Cheshire et al., 2008).

Position of YGM in utterances and turns

When examining the use of YGM in turns, we need to keep in mind that it is difficult to establish with any certainty whether the speaker intended to elicit a response. More precisely, a lack of response to the PM does not necessarily entail that a speaker-oriented function was intended. In the same vein, a rising intonation or the existence of a response does not necessarily entail that a hearer-oriented function was intended. Therefore, the analysis in this section will mainly focus on the following objective indicators:

a. discourse features that can be seen as prohibiting a response: utterance-initial and, more reliably, turn-initial uses;

b. the utterance- and turn-related positions of YGM uses that did receive a response including backchannels.

Table 11 shows the position of YGM in relation to the turn (initial, medial and final). In both corpora, the clear majority of uses are turn-medial, but the proportion is moderately (21.5%) higher in LIC3. The second most frequent position is turn-final, with MLEC2 speakers favouring this position almost twice (83.6%) as often as LIC3 speakers. It is also noteworthy that YGM was a full turn (or backchannel) in a small number of instances – as example (8) shows:

(8)   Roshan: he buns it down with man
        Robert: no
        Roshan: /you get me/
        Robert: / no no no .../
        Roshan: <laughs> ...

Table 11 Position in turn as initial, medial and final in MLEC2 and LIC3.

	MLEC2 (N=124)	LIC3 (N=141)	MLEC2 %	LIC3 %
T-initial	7	6	5.6	4.3
T-medial	76	105	61.3	74.5
T-final	36	25	29.0	17.7
Turn	5	5	4.0	3.5

The position of YGM within a turn is easy to identify, whereas its position in relation to an utterance is less straightforward. What guided the annotation of the latter was establishing the portion of the turn that YGM referred to in terms of content. For example in (9) below, the PM is in utterance-medial position, as it is wedged within the utterance it relates to. Table 12 presents the distribution of YGM in utterance initial, medial or final position.

(9)   Roshan: idiots {unclear} . they were still man that are a bit older but are cool they're .

         you get me . all friends and all that but

Table 12 Distribution of the position YGM in relation to the utterance in MLEC2 and LIC3.

	MLEC2 (N=124)	LIC3 (N=141)	MLEC2 %	LIC3 %
U-initial	10	7	8.1	5.0
U-medial	11	20	8.9	14.2
U-final	98	108	79.0	76.6
Utterance	5	5	4.0	3.5

In both corpora, almost 80% of uses are utterance-final. What is worth noticing is that the three other types of positioning (initial, medial, full utterance), accounting for more than a fifth of the instances, strongly suggest the speaker's intended pragmatic function. In these instances, the use is clearly speaker-oriented, as without a full utterance relating to the PM, the hearer cannot be expected to respond to its content. This expresses "subjectivity", because the use "index[es] speaker attitude or viewpoint" (Traugott, 2010, p. 32). It is also worth noticing that a sub-set of the utterance-initial uses – in fact, their majority – are also turn-initial. Further clues regarding the intended pragmatic functions of YGM can be derived from the examination of the frequency of cases in which YGM received a response, as well as its turn and utterance position in these instances.

Responses to YGM were a small minority in both corpora: 13.7% in MLEC2 and 7.1% in LIC3. This suggests that, whatever the speaker's intended pragmatic function, speaker- or hearer-oriented, it is usually treated by hearers as the former. Table 13 shows the proportion of responses received according to the position of YGM in relation to the utterance and turn.

Table 13 Proportion of responses in relation to position of YGM in MLEC2 and LIC3.

	MLEC2 (N=17)	LIC3 (N=10)	MLEC2 %	LIC3 %
U-medial + T-medial	1	0	5.9	0
U-final + T-medial	1	4	5.9	40.0
U-final + T-final	15	5	88.2	50.0
Utterance + Turn	0	1	0	10.0

In both corpora, most responses were received when YGM was both utterance-final and turn-final: MLEC2 (88.2%), LIC3 (50%). Further indications of the preferred pragmatic function of YGM are derived by examining the proportion of responses in the two positions that warrant a response, or, at least, are most likely to elicit one: utterance-final and turn-final (Table 14).

Table 14 Proportion of responses to YGM in utterance- and turn-final positions.

	MLEC2	LIC3
Utterance-final	15.3%	4.6%
Turn-final	41.7%	20%

A first observation is that a turn-final position seems more likely than an utterance-final one to receive a response: 41.7% of turn-final uses in MLEC2 and 20% in LIC3 received a response, whereas the proportion of responses after an utterance-final position were much smaller, as shown in Table 13. However, even in turn-final position, only a minority of YGM uses received a response, which further suggests that the dominant (perceived or intended) function of YGM is speaker-oriented.

Conclusion

The data presented in this chapter give new insights into how language forms are spreading in the community: *you get me* is currently used significantly more frequently by male non-Anglo speakers, but as we have shown, it is being adopted by other speakers. It is also used predominantly by male non-Anglo speakers outside of Hackney, and used more by teenagers

than younger speakers, which suggests that it is acquired within peer groups rather than at home, in line with other MLE features (Cheshire et al., 2011).

We have shown how *you get me* is positioned in utterances and turns. Few instances receive a response, demonstrating that the pragmatic function of YGM is largely speaker-oriented or reflecting subjectivity (Traugott, 2010). However, pragmatic markers may change over time from being largely expressions of subjectivity to expressions that reflect intersubjectivity, in that they become interactive elements receiving a response from the interlocutor, and are used to maintain flow in conversation (Fitzmaurice, 2004). In our data, this use of *you get me* only occurred in utterance-final or turn-final positions.

We have also shown that regression analysis, used predominantly in sociolinguistic studies, is not the only methodological option, and that the combined metrics of normalised frequency and spread can provide useful indications of emerging patterns, particularly when supported by the examination of the visual depiction of their interacting values.

Our investigation also shows that an increased awareness of corpus linguistic analysis methods is very timely in sociolinguistics: many older recordings are today being fully transcribed for the first time, such as a number of sociolinguistic interviews carried out in Philadelphia in the 1970s and 1980s (Labov, Rosenfelder, & Fruehwald, 2013). These analyses would have been impossible without very accurate and consistent orthographic transcriptions. Our conversion of two datasets originally conceived for traditional sociolinguistic analysis shows how such datasets can be examined using a corpus linguistic approach. We hope that our way of combining sociolinguistic data and corpus linguistic techniques can also lead to new insights from the analysis of other types of transcribed spoken data.

References

Anthony, L. (2014). AntConc (Version 3.4.3). Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp

Cheshire, J., Kerswill, P., & Williams, A. (2005). Phonology, grammar, and discourse in dialect convergence. In P. Auer, F. Hinskens, P. Kerswill (Eds.), *Dialect change: Convergence and divergence in European languages* (pp. 135–167). Cambridge: Cambridge University Press.

Cheshire, J., Fox, S., Kerswill, P., & Torgersen, E. (2008). Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. *Sociolinguistica, 22*, 1–23.

Cheshire, J., & Fox, S. (2009). Was/were variation: A perspective from London. *Language Variation and Change, 21*, 1–38.

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics, 15*, 151–196.

Cheshire, J., Adger, D., & Fox, S. (2013). Relative who and the actuation problem. *Lingua, 126,* 51–77.

Fitzmaurice, S. (2004). Subjectivity, intersubjectivity and the historical construction of interlocutor stance: from stance markers to discourse markers. *Discourse Studies, 6,* 427–448.

Foolen, A. (2012). Pragmatic markers in a sociopragmatic perspective. In G. Andersen & K. Ajmer (Eds.), *Pragmatics of Society* (pp. 217–242). Berlin: De Gruyter Mouton.

Fox, S., & Torgersen, E. (forthcoming). In N. Braber & S. Jansen (Eds.), *Sociolinguistics in England*. Basingstoke: Palgrave Macmillan.

Gabrielatos, C., Torgersen, E., Hoffmann, S., & Fox, S. (2010). A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics, 38,* 297–334.

Kerswill, P., & Williams, A. (2000). Creating a new town koine: children and language change in Milton Keynes. *Language in Society, 29,* 65–115.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2007). *Linguistic Innovators: The English of Adolescents in London: Full Research Report ESRC End of Award Report, RES-000-23-0680*. Swindon: ESRC.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2011). *Multicultural London English: The Emergence, Acquisition and Diffusion of a New Variety: Full Research Report ESRC End of Award Report, RES-062-23-0814*. Swindon: ESRC.

Labov, W. (1984). Field Methods of the Project in Linguistic Change and Variation. In J. Baugh & J. Sherzer (Eds.), *Language in Use* (pp. 28–53). Englewood Cliffs, N.J.: Prentice-Hall.

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language, 89,* 30–65.

Lehmann, C. (1985). Grammaticalization: synchronic variation and diachronic change. *Lingua e Stile, 20,* 303–318.

Maschler, Y. (1994). Metalanguaging and discourse markers in bilingual conversation. *Language in Society, 23,* 325–366.

McEnery, T. & Xiao, Z. (2004). Swearing in modern British English: The case of *fuck* in the BNC. *Language and Literature, 13,* 235–268.

Milroy, L. (1987). *Observing & Analysing Natural Language*. Oxford: Blackwell.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge
University Press.

Office for National Statistics. (2016, September 24). 2011 Census. Retrieved from
https://www.ons.gov.uk/census/2011census

Rayson, P., Leech, G. & Hodges, M. (1997). Social differentiation in the use of English
vocabulary: Some analyses of the conversational component of the British National
Corpus. *International Journal of Corpus Linguistics, 2,* 133–152.

Sankoff, D. (1988). Variable rules. In U. Ammon, N. Dittman, K. J. Mattheier, & P. Trudgill
(Eds.), *Sociolinguistics* (pp. 1150–1162). Berlin: De Gruyter.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Scott, M. (2016). WordSmith Tools version 7. Stroud: Lexical Analysis Software.

Sinclair, J. (1995). From theory to practice. In G. Leech, G. Myers, & J. Thomas (Eds.),
Spoken English on Computer (pp. 99–112). Harlow: Longman.

Stenström, A–B., Andersen, G., & Hasund, I. K. (2002). *Trends in Teenage Talk*. Amsterdam:
Benjamins.

Tagliamonte, S. A. (1998). Was/were variation across the generations: View from the city of
York. *Language Variation and Change, 10,* 153–191.

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

Tagliamonte, S. A., Smith, J., & Lawrence, H. (2005). No taming the vernacular! Insights
from the relatives in northern Britain. *Language Variation and Change, 17,* 75–112.

Torgersen, E., Gabrielatos, C., Hoffmann, S. & Fox, S. (2011). A corpus-based study of
pragmatic markers in London English. *Corpus Linguistics and Linguistic Theory, 7,* 93–
118.

Torgersen, E. & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua, 122,* 822–840.

Tormei, R. (2015). *Jamaican Speech Forms in Ethiopia*. Newcastle: Cambridge Scholars Publishing.

Traugott, E. C. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. In K. Davidse, L. Vandelanotte, & H. Cuyckens (Eds.), *Subjectification, Intersubjectification and Grammaticalization* (pp. 29–74). Berlin: De Gruyter.

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.