Using social media to infer the diffusion of an urban contact dialect

DOI: 10.1111/josl.12653

ARTICLE

Journal of Sociolinguistics WILEY

Using social media to infer the diffusion of an urban contact dialect: A case study of Multicultural London English

Christian Ilbury[1] | **Jack Grieve**[2] | **David Hall**[3]

[1]Department of Linguistics and English Language, The University of Edinburgh, Edinburgh, UK

[2]Department of English Language and Linguistics, The University of Birmingham, Birmingham, UK

[3]Independent Researcher

Correspondence
Christian Ilbury, Department of Linguistics and English Language, The University of Edinburgh, Edinburgh EH8 9AD, UK.
Email: cilbury@ed.ac.uk

Funding information
Arts and Humanities Research Council (UK); the Economic and Social Research Council (UK); Jisc (UK), Grant/Award Number: 3154; the Institute of Museum and Library Services (US); Digging into Data Challenge

Abstract

Sociolinguistic research has demonstrated that 'urban contact dialects' tend to diffuse beyond the speech communities in which they first emerge. However, no research has attempted to explore the distribution of these varieties across an entire nation nor isolate the social mechanisms that propel their spread. In this paper, we use a corpus of 1.8 billion geo-tagged tweets to explore the spread of Multicultural London English (MLE) lexis across the United Kingdom. We find evidence for the diffusion of MLE lexis from East and North London into other ethnically and culturally diverse urban centres across England, particularly those in the South (e.g. Luton), but find lower frequencies of MLE lexis in the North of England (e.g. Manchester), and in Scotland and Wales. Concluding, we emphasise the role of demographic similarity in the diffusion of linguistic innovations by demonstrating that this variety originated in London and diffused into other urban areas in England through the social networks of Black and Asian users.

KEYWORDS

diffusion, Multicultural British English, Multicultural London English, multiethnolects, social media, Twitter, urban contact dialects

1 | INTRODUCTION

Over the past two decades, research has increasingly documented a new variety of British English—what has been termed 'Multicultural London English' (MLE)[1] (Cheshire et al., 2008, 2011). Although this variety was first studied in East London, more recent work has identified similarities between MLE and varieties spoken in other urban areas of England, including the cities of Manchester and Birmingham. This has led some to argue for a more general 'Multicultural English' (ME: Fox et al., 2011; Khan, 2006) or 'Multicultural (Urban) British English' (MBE: Drummond, 2018), defined as an 'overarching variety or repertoire of shared features, with each urban centre then having its own local version or sub-variety' (Drummond, 2018: 12).

The emergence of MLE and ME/MBE can be seen as an example of a more general phenomenon: the appearance of new 'urban contact dialects' (Kerswill & Wiese, 2022). There is now a wealth of research which describes the development of these varieties in countries such as Sweden (Gross & Boyd, 2022; Kotsinas, 1988), Denmark (Aasheim, 1997; Quist, 2008, 2022), Norway (Svendsen, 2022; Svendsen & Røyneland, 2008), the Netherlands (Nortier, 2008; Kossmann, 2017), Germany (Şimşek & Wiese, 2022; Wiese, 2009), and France (Cheshire & Gardner-Chloros, 2018; Gadet, 2022). Some of this work has also described the tendency for urban contact dialects to spread (or diffuse) beyond the speech communities in which they first emerge (e.g. Dirim, 2005; Christensen, 2012). However, because this research focusses on spoken language recordings from individual speech communities, we know very little about the extent of their spread and the social mechanisms that propel their diffusion. Although it has often been assumed that these varieties spread from urban centre to urban centre (see Meyerhoff, 2022)—in a way compatible with Trudgill's (1974) formulation of the 'gravity model' of linguistic innovations—no research has examined the diffusion of an urban contact dialect across an entire nation.

For reasons of practicality and time, these issues are unlikely to be addressed using traditional sociolinguistic methods. Therefore, in this paper, we harness the analytical potentials of big data to shed light on the diffusion of an urban contact variety, concentrating on the spread of MLE across the United Kingdom and the actuation of MBE. Focussing on lexical variation in Twitter[2] data, we ask three main research questions:

1. What is the overall regional distribution of MLE lexis across the United Kingdom?
2. What can this distribution tell us about the origins of MLE?
3. What can this distribution tell us about the diffusion of MLE?

To answer these questions, we analyse the frequency of MLE lexis in a dataset of 1.8 billion geo-located tweets. By mapping tweets containing MLE lexis, we identify areas where the variety is used most frequently. Assuming that the regional distribution of words reflects where the dialect is used, we find compelling evidence to support the claim that MLE originated in East and North London (see Cheshire et al., 2008, 2011; Fox, 2015), and that MLE lexis diffused out from London primarily into other urban and ethnically diverse areas of England but not across Britain as a whole. In fact, we find lower rates of MLE lexis in the North of England, most notably in Manchester (cf. Drummond, 2018). We also find low frequencies of MLE lexis in Scotland, Northern Ireland, and most of Wales. We

[1] The labels we use in this paper (MLE/MBE) are academic terms. Although 'MLE' is now common in media, speakers of the variety often refer to MLE as 'slang' (Kerswill, 2013: 149; Ilbury, 2019; though see Walcott, 2022).

[2] In July 2023, Twitter was renamed 'X'. We refer to this platform as 'Twitter' throughout given that our data were collected prior to this rebrand.

therefore question the extent to which this variety could be truly considered a Multicultural *British* English.

Further, by examining the trajectory of the diffusion, we explore the social mechanisms behind the spread of MLE lexis and, by extension, the variety. We argue that the diffusion is not explained simply by distance from London alone, as we might predict on the basis of the *Wave Model* of diffusion (Schmidt, 1872), nor is it explained by distance and population density, as predicted by the *Gravity Model* (Trudgill, 1974) of linguistic change. Rather, we show that it is the ethnic and social diversity of the urban centre and, relatedly, the shared friendship, familial and cultural ties of users across cities that predicts the spread. Indeed, the lexical repertoire of MLE we identify is not evenly distributed across England—or even in London. Rather, MLE lexis is concentrated in areas that are relatively close to London, have greater numbers of lower socio-economic status (SES) households, and where there are sizeable numbers of people identifying as Black or Asian. Our findings not only provide evidence for the role of demographic similarity in the spread of linguistic innovations in-line with other recent research in Computational Sociolinguistics (CS; Eisenstein et al., 2014; Grieve et al., 2018) but also demonstrate the potential of big data in exploring the development and diffusion of new urban contact dialects.

2 | BACKGROUND

2.1 | Multicultural London English

First described in the 'Linguistic Innovators' project (2004–2006), MLE was initially defined as a new sociolect of English that was spoken mainly by young working-class people living and working in inner city neighbourhoods in London (Cheshire et al., 2008, 2011; Fox, 2015). In earlier work, MLE was defined as a 'multiethnolect' (Clyne, 2000) based on the finding that speakers of different ethnic backgrounds variably used features of the variety. However, more recently, the concept of the 'multiethnolect' has become contested (see Jaspers, 2008), including in work on MLE (Ilbury & Kerswill, 2024). Following Wiese (2022), we instead refer to MLE as a new 'urban contact dialect'. When defined in these terms, MLE can be considered a variety of English that 'emerged in contexts of migration-based linguistic diversity among locally born young people, [which is used to mark] speakers as belonging to a multiethnic peer group' (Wiese, 2022: 117).

MLE appears to have developed during the post-war period when there was considerable population movement in East London. After WWII, many White working-class residents (i.e. Cockneys) moved out into neighbouring counties such as Essex whilst, at the same time, following a call for increased migration to fill post-war labour shortages, migrants from Commonwealth states (e.g. Jamaica, India) settled in the area. East London is now home to sizeable migrant communities from a diverse range of countries, including Jamaica, Cyprus, Vietnam, and Bangladesh (Hackney, 2016).

Today, MLE has largely supplanted the traditional working-class vernacular, Cockney English. There are several distinctive and innovative linguistic features of MLE that are not present in other varieties of London English (see Cheshire et al., 2008, 2011 for an overview). MLE is perhaps most differentiated from other London varieties in terms of its phonology. A common phonological feature of MLE is that some diphthongs tend to be near-monophthongs, hence *face* becomes [fes], and *lie* is often [laː]. Grammatical and discourse features include the pronoun *man* as in 'man said "one two nine"' (Cheshire, 2013; Hall, 2020; Ilbury, 2019) and sentence final *still* as in 'you've got a car though, still' (Cheshire et al., m.s.). Finally, and most relevant to our analyses, MLE is associated with

an abundance of new lexis. Words used by MLE speakers include *leng* 'nice, fit, attractive' or (less frequently) 'gun', *blud* 'mate, brother' and *paigon* 'a deceitful person'.

Many features of MLE have been heavily influenced by, if not borrowed directly, from Jamaican English (henceforth JE) and Jamaican Creole/Patwah (or 'Patois'). This is because London (particularly, the South and East, for example Brixton and Hackney) is home to large diasporic Caribbean communities who have contributed significantly to the culture and language of the city. The influence of Caribbean varieties on MLE can be seen at multiple linguistic levels, but it is particularly apparent in the lexicon. Lexis and phases, such as *yard/yaad* 'house' and *wha gwaan* 'what's going on', originally from JE are now common in MLE.

Nevertheless, although MLE is heavily influenced by JE and is often described as a 'Black British vernacular' in media and popular discourse (see Ilbury & Kerswill, 2024; Walcott, 2022), earlier research on MLE claimed that it was an 'ethnically neutral variety' (Cheshire et al., 2011: 157). Rather than speaker ethnicity, Cheshire et al. (2008) found that the ethnic diversity of the individuals' friendship network predicted the use of MLE, with speakers with more ethnically diverse friendship networks being more likely to use MLE features than their peers with ethnically homogenous networks. Consequently, people from all backgrounds can now be observed to variably use features of MLE.

The association between MLE and Blackness is therefore likely to be indicative of a type of 'raciolinguistic enregisterment'—a process by which linguistic features become 'emblematized as sets of signs that correspond to racial categories' (Rosa & Flores, 2017: 632). Indeed, MLE/MBE speakers are often perceived as 'sounding Black' (Drummond, 2016) and in earlier media accounts of the variety, this association was often made explicit when MLE was labelled 'Jaifaican' (literally 'fake Jamaican'). At least at the level of lexis, this ideological association is in part due to the very many lexical borrowings from Caribbean Englishes.

More recent evidence for the raciolinguistic enregisterment of MLE is found in popular media accounts of the variety which describe it as a 'black, inner city language' (Hirsch, 2018: 7) and a 'street parlance connected intimately to the black diaspora' (Boakye, 2019: 363). Research has also found that MLE is indexically associated with Black British identities through grime (Drummond, 2018)—a Black British electronic music genre that originated in East London in the early 2000s—and 'road culture'—a Black British interpretation of North American street cultures (Boakye, 2019; Ilbury, 2023).

2.2 | The diffusion of urban contact dialects

In addition to the previous developments, scholars have also documented similarities between MLE and varieties elsewhere in England, such as those spoken in Birmingham and Manchester (Drummond, 2018; Fox et al., 2011; Khan, 2006). For instance, in research on youth language in the North West city of Manchester, Drummond (2018) observed that many speakers had similar vowel inventories to those in the MLE project. They also appeared to use some other MLE features, such as TH-stopping and quotative *be like*, and were recorded using *rah, peng, peak* and other lexis associated with MLE. This leads Drummond to argue for the existence of 'MBE' which he defines as a more general variety that incorporates features of MLE and the local dialect. Comparable arguments are made by Fox et al. (2011) who proposed the term 'Multicultural English' (ME) based on similarities between the speech of young people in London and Birmingham. In this study, we adopt Drummond's (2018) term 'Multicultural British English' (MBE) to refer to a possible general urban contact dialect of British English.

Arguably, the development of MBE is indicative of a more general phenomenon in which new urban contact dialects diffuse beyond the speech communities in which they first emerge. In Germany, Kiezdeutsch—a variety originally associated with the Kreuzberg district of Berlin—is now used elsewhere such as in the city of Hamburg (Dirim, 2005), whereas work on the Copenhagen variety has documented its use in Århus—a city on the Jutland peninsula's east coast (Christensen, 2012). To date, however, most of this work has focussed on describing the use of these varieties based on recordings of a limited number of speakers from individual speech communities. As such, researchers have not been able to examine the diffusion of these varieties across multiple regions nor isolate the social mechanisms that propel their spread (though see Grondelaers & Marzo, 2023). Although there is some reason to assume that 'urban varieties jump from major urban centre to major urban centre' (Meyerhoff, 2022: 153), this conclusion has not been confirmed by empirical data. Our understanding of the development and spread of urban contact dialects is therefore limited by a methodological inability to examine their use across *multiple* locations and regions.

These limitations are particularly apparent in the work on MLE. First, it is unclear whether MLE did actually originate in East London or whether the identification of this variety here is due to the overrepresentation of sociolinguistic research in the area (see *inter alia* Cheshire et al., 2008, 2011; Fox, 2015; Gates, 2018; Ilbury, 2019). Indeed, researchers have documented MLE in other areas of the city, such as the West London borough of Ealing (e.g. Oxbury, 2021; Oxbury & McCarthy, 2019). Second, the status of MBE and its relationship to MLE remains unclear. Although there is some reason to suspect a priori that MBE is a development of MLE, there is no empirical evidence on the matter and the issue is still contested. Drummond (2021), for instance, claimed that MBE 'arguably' emerged from MLE, whereas Fox et al. (2011) appeared to suggest that the comparable demographics of London and Birmingham led to simultaneous (but related) developments. In the current paper, we attempt to resolve these issues by harnessing big data to track the spread of MLE across multiple areas of the United Kingdom. In doing so, we emphasise the potential of this approach in informing our understanding of the development of urban contact dialects more generally.

3 | METHODS

3.1 | Corpus

The corpus we analyse in this paper comprises 180 million geolocated tweets totalling 1.8 billion words that were downloaded using the Twitter API over the span of 1 year, from 1 Jan 2014 to 31 December 2014 (see Grieve et al., 2019 for a description of the dataset). The dataset is drawn from 1.9 million unique accounts, with a median of 10 tweets per user. Tweets were collected over a period of 360 days, with data for 5 days missing due to technical issues. The corpus is not filtered in any way, such that retweets and spam posts are not removed. As such, the dataset is representative of the content that users see when using the platform.

All tweets are geolocated with the precise longitude and the latitude of the user when they posted the message using a GPS-enabled smart phone. This allows us to sort all tweets into 124 postal code areas, effectively stratifying the corpus into 124 regional sub-corpora, which can then be used as the basis of mapping and spatial analysis. On average, the corpus contains 1.5 million tweets per region. The number of tweets varies from region to region, from 5.5 million tweets in Manchester to 54,000 tweets in the Outer Hebrides. Notably, London is divided into several smaller postal code regions. We use postal region data to enable comparison with other analyses of UK Twitter data which have used a similar approach (e.g. Grieve et al., 2019).

TABLE 1 The list of 75 independent Multicultural London English (MLE) associated words.

1. alie	26. drag	51. prang
2. bare	27. duppy	52. raasclat
3. batty bwoi/boi/boy	28. endz	53. rah
4. batty	29. fam	54. roadman
5. beanie	30. famalam	55. rude boi
6. beef	31. gassed	56. rudeboy
7. blud	32. glowed up	57. safe
8. bludclaart	33. gully	58. shot
9. bomboclaat	34. gullyside	59. shotta
10. booky/booukie/bookie	35. gyal	60. shotter
11. bora	36. gyaldem	61. sket
12. boydem	37. hench	62. sket
13. bredrin	38. jokes	63. slew
14. breeze (Adj)	39. jook,	64. swag
15. breh	40. leng	65. tekkers
16. bruck	41. lips	66. ting
17. bruk	42. long	67. uck
18. bun (V)	43. mandem	68. wasteboy/girl
19. butters	44. manor	69. wasteman/girl
20. ching	45. merk	70. whagwaan
21. chirpse	46. murk	71. yard
22. clapped	47. nang	72. yat
23. cotch	48. nitty	73. youngers
24. deadting	49. olders	74. yout
25. dench	50. paigon	75. yute

3.2 | Lexical analysis

To analyse the geographical spread of MLE lexis, we first developed a list of 75 independent words that are associated with MLE (Table 1). The list comprises the following:

1. Home-grown MLE terms, such as *leng* 'attractive', *paigon* 'enemy/untrustworthy person', and *uck* 'fellatio'.
2. Words from Caribbean Englishes, such as *bludclaart* 'period pad' (used as an insult or intensifier), *yard* 'home', *rah* (an exclamation).
3. And words which are often considered part of a more general youth style, such as *butters* 'ugly', *jokes* 'funny', *long* 'tiresome/annoying'. These words may have originated in MLE (or at least in London) but have now entered general circulation and are used more regularly by speakers who do not use other features of MLE/MBE (including by Ilbury and Hall).

The list we develop is based primarily on previous research on the variety (e.g. Cheshire et al., 2008, 2011) including our own experience of conducting fieldwork on MLE and other London varieties (e.g. Hall, 2020; Ilbury, 2019, 2023), and our (Ilbury & Hall) emic insights as Londoners. We

TABLE 2 The list of 47 core Multicultural London English (MLE) words.

1. alie	13. ching	25. gyaldem	37. shotta
2. bare	14. chirpse	26. hench	38. sket
3. batty	15. clapped	27. leng	39. slew
4. beef	16. cotch	28. mandem	40. tekkers
5. beanie	17. dench	29. nang	41. ting
6. blud	18. duppy	30. nitty	42. uck
7. booky	19. endz	31. paigon	43. wasteman
8. bora	20. fam	32. prang	44. yard
9. bredrin	21. famalam	33. rah	45. yat
10. breh	22. gassed	34. roadman	46. yout
11. bruk	23. gully	35. rude boi	47. yute
12. butters	24. gyal	36. rudeboy	

also consulted social media platforms (TikTok, Instagram) and the crowd-sourced online dictionary for 'slang', urbandictionary.com, where these terms are often described as 'MLE' or 'London slang'. Nevertheless, we acknowledge that classifying lexical features into discrete varieties is not entirely viable, and although we base this list on the previous research we are, in essence, contributing to the (raciolinguistic) enregisterment (Agha, 2007; Rosa & Flores, 2017) of MLE. However, our goal is not to define such varieties. Rather, we use these categories as a way to ensure that we consider as wide a range of words related to MLE as possible in order to maximise our chances of identifying patterns of MLE Lexis from the ground up, through a multivariate statistical analysis as presented in the next section, rather than to constrain our analysis or to make assumptions about what types of patterns we would find.

In the first part of our analysis, we measure the individual frequency of each of the putative MLE lexical items across the 124 postal code regions and map the results. Given that the number of tweets varies among different regions, the data were first normalised, in this case per million words. Words with no tokens at all—that is a word that does not occur in a given postcode region—were removed.

In the second half of our analysis, we develop an aggregated map based on a more restricted list of 47 MLE words (see Table 2) which we call 'core MLE' words. We focus on these words for several reasons. First, these words are often considered typical of MLE in that they are frequently used by speakers in naturalistic spoken recordings (e.g. Cheshire et al., 2008, 2011; Ilbury, 2019). Second, these words are not used in other (unrelated) dialects. For this reason, we did not include words like <merk> which in MLE means to 'stab/kill', since <merk> is used in Scots to refer to a long-obsolete Scottish silver coin. This list is, in part, based on the patterns identified in the first part of our analysis where we focus on the distribution of the 75 independent words.

To identify common patterns of regional variation in MLE lexis, we subjected the full set of 47-words to a principal component analysis (PCA) after scaling the variables, so that each word has the potential to contribute equally to the analysis, rather than just the most frequent forms (see Grieve et al., 2011). The PCA analysis essentially compares the similarity of the maps for every pair of words, generating a series of aggregated dimensions that identify the most common regional patterns instantiated in those maps, ranked by order of importance. Each of these aggregated dimensions is associated with a set of *dimension loadings* for each word, which reveal how strongly the map for each word is represented by that dimension, and a set of *dimension scores* for each location, which can be mapped to visualise that underlying pattern of regional lexical variation. In this way, we are able to identify areas where MLE

lexis, or possibly areas where different types of MLE lexis, are used most or less frequently, allowing us to better understand how the large number of diverse lexical items we are analysing are related to one another and to identify regions that are most strongly associated with the frequent usage of these forms. In essence, this approach allows us to uncover areas where MLE clusters as a coherent repertoire of lexical features (Cheshire et al., 2008).

To infer the social identities of authors and the demographics of an area, we compared the distribution of MLE words with data from the 2011 UK population census (ONS, 2011). This gives us a broad snapshot of the demographic composition of the United Kingdom at a time point closest to when our data were collected (2014). The census provides information on various social measures, including the age, sex, and ethnic composition of a given area. These data are largely derived from self-selected responses to the census. For instance, participants were asked to identify their 'ethnicity' by self-selecting from labels which included 'White—English/Welsh/Scottish/Northern Irish/British', 'Asian/Asian British—Indian', and 'Mixed/Multiple ethnic groups—White and Asian'. We acknowledge that by using census categories as analytical frames, we run the risk of homogenising the inherent (linguistic) variation *within* social groups (see Blake, 2014). However, we contend that broad-level ethnic and racial labels (e.g. Black British) *do* often become meaningful identities for individuals and communities. Relevant to the current paper, Boakye (2019) has suggested that the label 'Black British' has become a 'tangible' identity for very many people today. This is evidenced by the designation of Grime music as a Black British cultural innovation as well as the apparent raciolinguistic enregisterment of MLE and its proposed status as a 'Black British vernacular' (see Boakye, 2019; Drummond, 2018; Hirsch, 2018; Ilbury, 2023; Walcott, 2022). Thus, although we do not wish to suggest that any social group is linguistically or culturally monolithic, we believe that using census information is the best approach available to us in identifying the broad social distribution of MLE. This approach is also comparable to other CS analyses (e.g. Eisenstein, 2015).

3.3 | Limitations

Before presenting our analysis, we should acknowledge some additional limitations of our data and method. The first issue concerns the time period represented by the corpus, which is now a decade old. The corpus was compiled before Twitter introduced limitations on the streaming API. Such limitations have heavily restricted the amount of data that can be extracted from the platform such that a corpus of this size could not be feasibly compiled using these methods today. We maintain that this corpus is an extremely valuable and unique resource in answering the research questions we set out to explore.

In fact, though the age of the dataset could be problematic for other analyses, the time period in which the data were collected is arguably advantageous given the scope of our research questions. Recently, MBE/MLE has become synonymous with the music genre of grime and is often used as a more general youth style, such that features of the variety are often used stylistically by speakers to index their engagement with these social practices and communities (see Drummond, 2018; Ilbury, 2019, 2023). This is even apparent outside of the United Kingdom, where MLE lexis appears to be used by individuals from geographically disparate countries such as Australia and Spain who listen to grime music (The Economist, 2021). The year in which our data were collected, however, predates the mainstream adoption of MLE. In 2014, grime music was, arguably, still a fringe subculture (Collins, 2014). The time period of this dataset therefore allows us explore the spread of MLE at a comparable time point to other research on the use of MLE outside of London (Fox et al., 2011; Khan, 2006) and before the mainstreaming of MLE through grime music (cf. Drummond, 2018).

The second limitation of this approach is polysemy. Many MLE terms have Standardised English (SE) equivalents with very different meanings. For instance, *peak* refers to 'the highest point' in SE, whereas in MLE it is used as an adjective meaning 'bad, shame'. In a dataset of this size, however, it is not feasible to distinguish between all possible senses of a word. Although this may initially appear to be a significant limitation of exploring lexical variation in Twitter data, we show later in Section 4.2 that polysemy does not have any substantial effect on the patterns we identify.

Finally, we should acknowledge the potential issues of extrapolating the trends we identify based on written data (i.e. Twitter) to the status of a variety of English that is typically used in spoken interaction (cf. Cheshire et al., 2008, 2011). Here, we follow other (CS) researchers who have argued that the linguistic analysis of social media data is not only intrinsically valuable, given the importance of this modern form of communication, but can be a useful proxy for spoken language in studies of language variation and change (see *inter alia* Grieve et al., 2017, 2018; Huang et al., 2016; Ilbury, 2020). Research has consistently demonstrated that linguistic variables in social media texts (mostly Twitter) are often subject to similar social and linguistic constraints as the corresponding variable in speech. For instance, in an analysis of variation on Twitter, Eisenstein (2015) found that orthographic representations of the phonological variables (ing) and (t,d) are subject to the same social and stylistic constraints as the spoken variables, whereas Grieve et al. (2019) observed a remarkable similarity in the geographical distribution of UK dialect lexis in Twitter data when compared with traditional dialect maps from the BBC Voices dialect survey. We therefore assume that variation in social media texts (in this case lexis) can serve as a useful proxy for speech, and that the patterns we identify in our Twitter dataset can reliably be used to infer the development of a spoken language variety (i.e. MLE).

Nevertheless, it is important to remember that the results of this study, or of any analysis of Twitter data, can only at a maximum generalise directly to this specific variety of online communication. Although given the difficulties of exploring the spread of MLE across multiple geographical areas using traditional methods, we emphasise the analytical value and potential of social media data in exploring the geographical diffusion of urban contact dialects in data from millions of users across an entire nation.

4 | ANALYSIS

4.1 | Independent word analyses

In the first section of our analysis, we present maps that plot the relative frequencies of the 75 individual words that are associated with the MLE lexicon (see Table 1). Our assumption is that, to some extent, regional patterns in the relative frequencies of these forms at one point in time (2014), allows us to infer information about the spread of these individual words over time. We assume areas where these words are used more frequently are also areas where these words spread earlier, especially if there are clear regional patterns that are consistent with some kind of diffusion are visible.

We report the patterns for some of the 75 words here. Later, we restrict our analysis, in part justified by this analysis of individual words. We begin by considering the maps for some of the words that are often considered typical of MLE. Figure 1 maps the relative frequency of *paigon* ('enemy' or 'untrustworthy person', as in 'he's a paigon') per million words across the United Kingdom. The areas that are shaded dark red are 'hotspots' where there is a higher relative frequency of the word. As one can see, there is a high concentration of *paigon* in the South of England, particularly in London, as indicated in the inset of the city. In fact, *paigon* appears to be highly frequent in all regions of London, albeit less common in West and South West London. Outside London, we find some evidence for the northwards

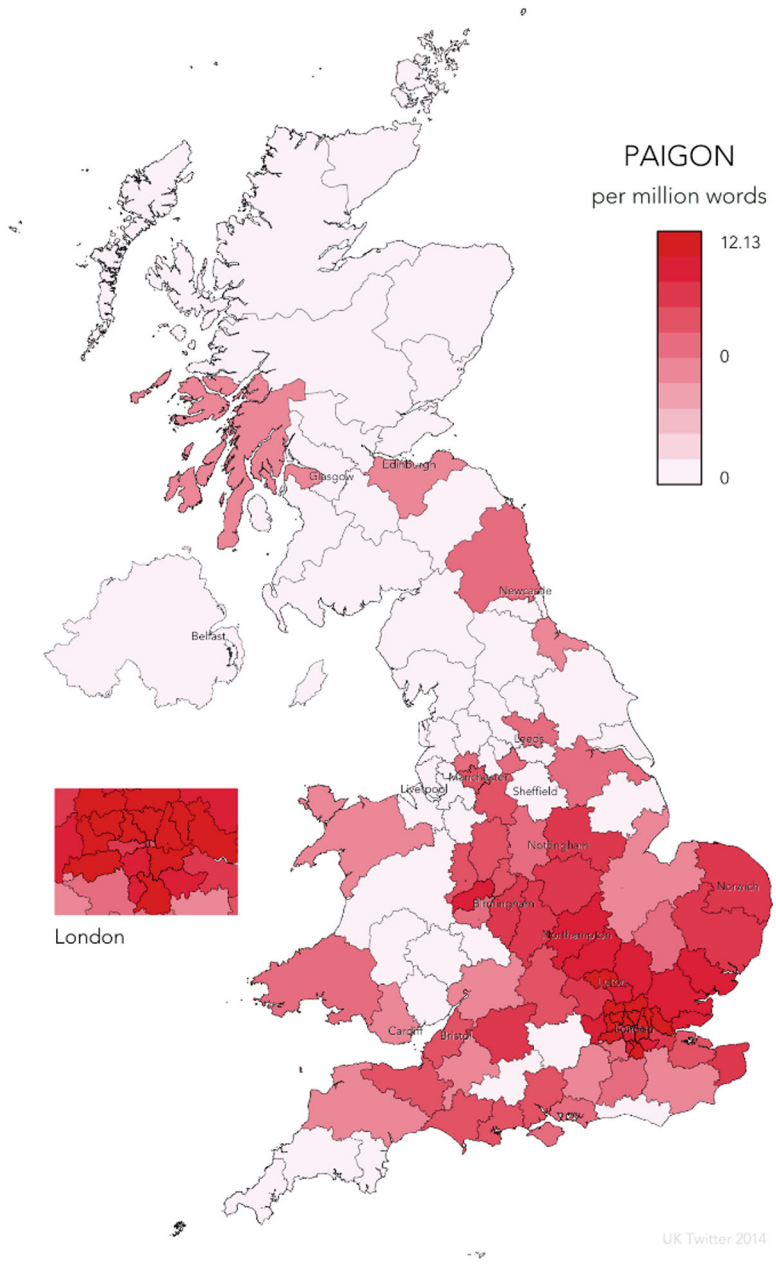

FIGURE 1 Frequency of 'paigon' across the United Kingdom per million words.

spread of *paigon* into cities such as Luton, Northampton, and Birmingham. All these areas are urban and ethnically diverse. For instance, although, comparably somewhat less diverse than London, Luton is home to large numbers of people identifying as 'British Asian' and 'Black British', comprising 30% and 9.8% of the population, respectively (ONS, 2011). Areas with very low frequencies of *paigon* are those which are both rural and geographically and culturally disparate from London. This includes the South West of England, Northern Ireland, and Scotland. When *paigon* is used in these regions, this word occurs only at very low rates and is largely restricted to major urban centres, such as Glasgow and Edinburgh.

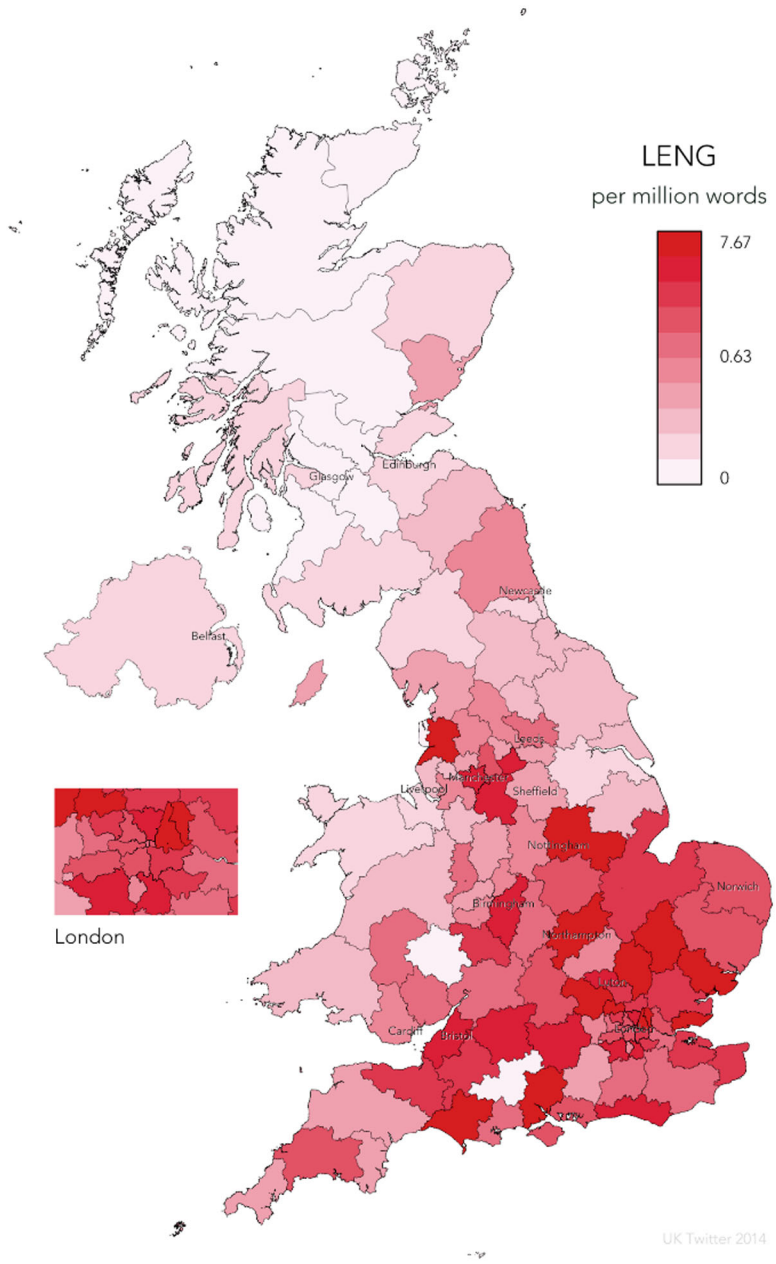

FIGURE 2 Frequency of 'leng' across the United Kingdom per million words.

This pattern is broadly consistent when we examine the geographical distribution of other words that we considered to be typical MLE words. Figure 2 shows the distribution of *leng* per million words—an adjective that is most often used to mean 'nice' or 'attractive' as in, 'that girl is leng'. As before, the darker red shading indicates a higher frequency of the word in that area. Similar to *paigon*, Figure 2 shows that *leng* is highly concentrated in the South East of England, particularly in London. The inset shows that there are somewhat higher frequencies of *leng* in the North and East of London. We again see evidence for the spread of MLE lexis into other urban areas such as Northampton, Birmingham,

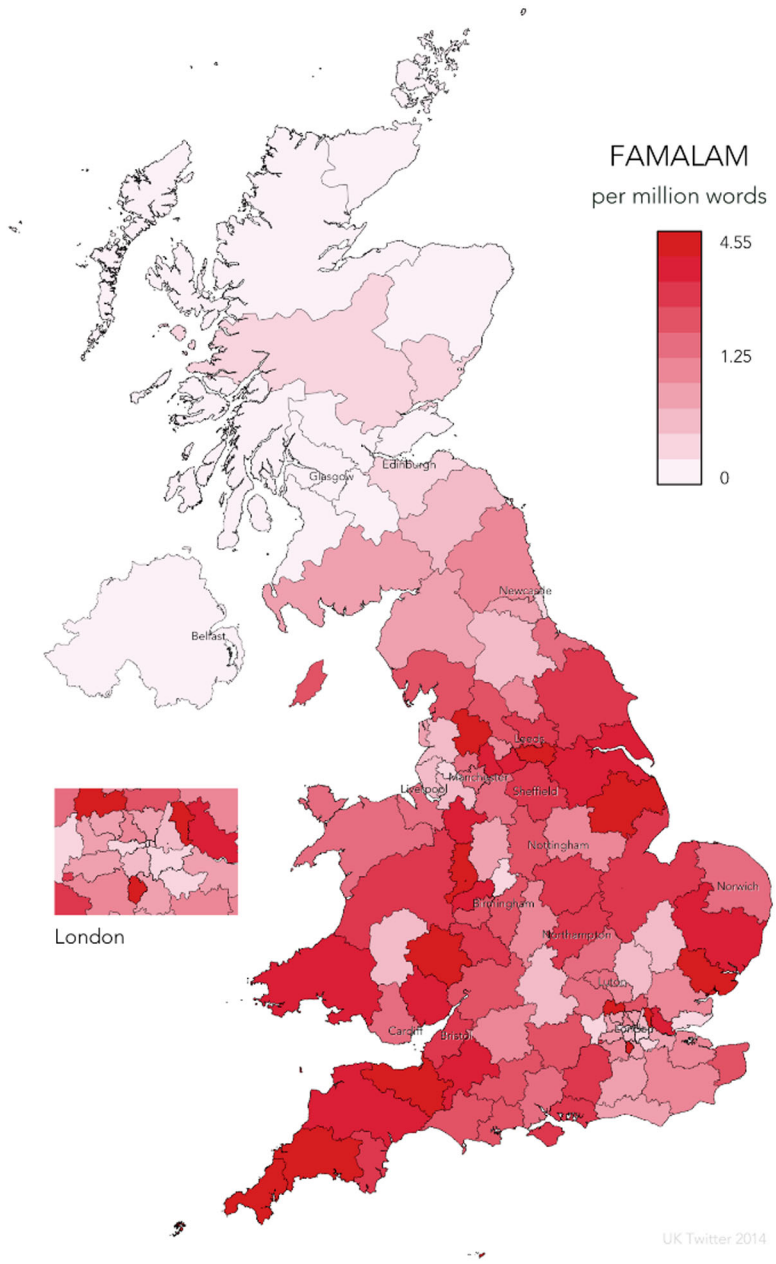

FIGURE 3 Frequency of 'famalam' across the United Kingdom per million words.

Nottingham, and Manchester, albeit at somewhat lower rates than London. Likewise, we see much lower frequencies of *leng* in areas that are distant from London, such as those in Wales, Northern Ireland, and Scotland.

Thus far, our analysis has considered two very typical MLE words, *leng* and *paigon*. However, some MLE words have become more widespread. This includes words such as *famalam* 'friend/family' as in 'what's going on, famalam?' and *butters* 'ugly' as in 'those shoes are butters'. Compared to *paigon* and *leng*, we see an altogether different pattern for the geographical spread of these words. Figure 3

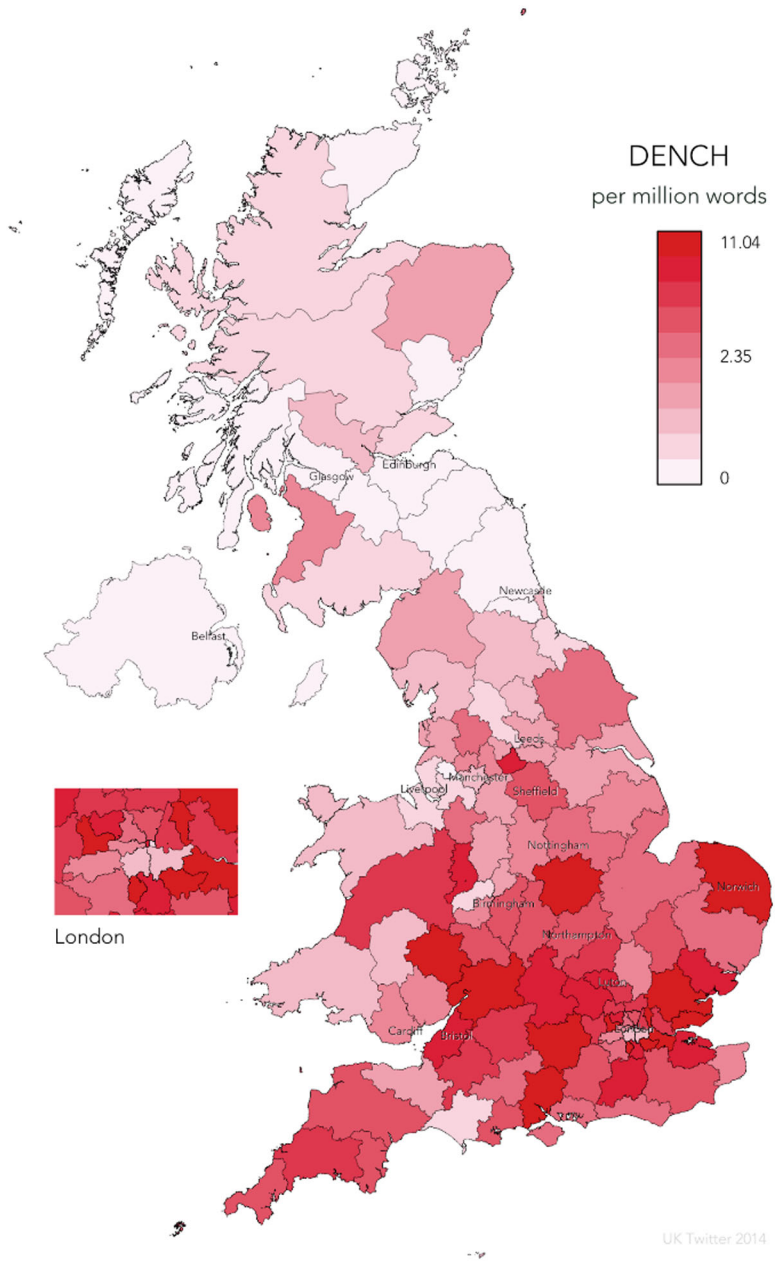

FIGURE 4 Frequency of 'dench' across the United Kingdom per million words.

shows that *famalam* is highly frequent in the South West and North West of England and parts of Wales. However, crucially, and unlike *paigon* and *leng*, it is relatively infrequent in London and most of the South East of England. This pattern is most apparent in the inset where we can observe very low frequencies of this word across London. *Famalam* is also used less frequently in regions outside of London where we previously observed relatively high rates of MLE lexis (e.g. Luton).

A similar pattern is seen for other words such as *dench* 'attractive or physically impressive' as in 'that guy is dench'. The term was popularised by the London-based grime artist, Lethal Bizzle, and has become somewhat of a stereotype of the variety. Figure 4 shows the distribution of *dench* across the

United Kingdom. As in the case of *famalam*, we see that *dench* is used much less in London and the South East of England. Rather, hotspots emerge in geographically disparate areas in England, including around Norwich, Bristol and Southampton.

Given the somewhat inconsistent findings here, how can we explain the relative spread of lexis like *paigon* and *leng* with reference to *famalam* and *dench*? One plausible and very possible interpretation is that *famalam* and *dench* were some of the earlier features to have spread. Once they spread, and having been adopted into more general circulation, they were potentially dropped by MLE users. Thus, their association with MLE possibly exists only as a stereotype of the variety, with those using MLE as a habitual style, using these words much less frequently. Indeed, neither *famalam* nor *dench* are recorded by Ilbury (2019) in over 40 h of spoken interactions of MLE speakers. This contrasts with their use in popular culture where these words have attained some stereotypical or symbolic association with MLE, such as the BBC comedy 'Famalam' which first aired in 2017.

Given that the lexical inventory of MLE is heavily influenced by JE and other Caribbean varieties, we also explore the geographical distribution of MLE-associated terms that are borrowed from these varieties. Figure 5 maps the frequency of the JE/Patwah term *gyaldem* 'girls/women', as in 'look at the gyaldem' (see Cassidy & Le Page, 2002; Sebba, 1993).

As one can see, the distribution of *gyaldem* is very different to the MLE words previously discussed. Although we do still see high frequencies of *gyaldem* in London, Luton and Birmingham, we see that *gyaldem* is also used in areas where *paigon, leng* and other 'core' MLE words were not very frequent, such as Bristol and Leeds. Perhaps unsurprisingly, these are areas with relatively sizeable Black communities. In the 2011 census, 6% of residents identified as 'Black British' in Bristol, whilst in Leeds, 3.5% of residents identified as such. These are some of the largest Black communities outside London and the South East of England. In less ethnically diverse areas such as Norwich (1.6% Black British), we see much lower frequencies of *gyaldem* and other JE words. It is therefore likely that we see higher frequencies of JE lexis in these areas simply because these regions are home to relatively sizeable JE/Patwah speaking communities. Evidently, then, the presence of Black British communities alone does not predict the use of MLE lexis in an area.

4.2 | Aggregate analyses

In the previous section, we examined the geographical distribution of 75 individual words that were associated with MLE. Through this analysis we identified what appear to be distinct geographical patterns of lexis. However, although we have uncovered some patterns which look broadly similar, it is unclear what this more general MLE pattern looks like or how strongly this pattern is instantiated across the full feature set, including compared to any other common regional patterns that might also be present. To address these questions, we use a PCA to identify underlying dimensions of variation in the relative frequencies of the 47 'core' MLE words (Table 2) measured across the 124 UK postal code areas. The PCA effectively reduces the number of dimensions we use to describe variation in this dataset from the 47 individual words to a small number of distinct aggregated dimensions that account for as much of the distinct patterns of shared variance in the values of these words. In this way, a PCA makes the data more interpretable, allowing for common patterns to be identified, while preserving variability in the original data. The PCA also means that we can identify regions which are highly correlated with those patterns by mapping each of these aggregated dimensions, thus detecting regions where MLE lexis is particularly frequent.

Despite the ability of a PCA to identify multiple independent dimensions of variation in multivariate data, each of substantial importance, in this case, the PCA remarkably finds that there is only *one main*

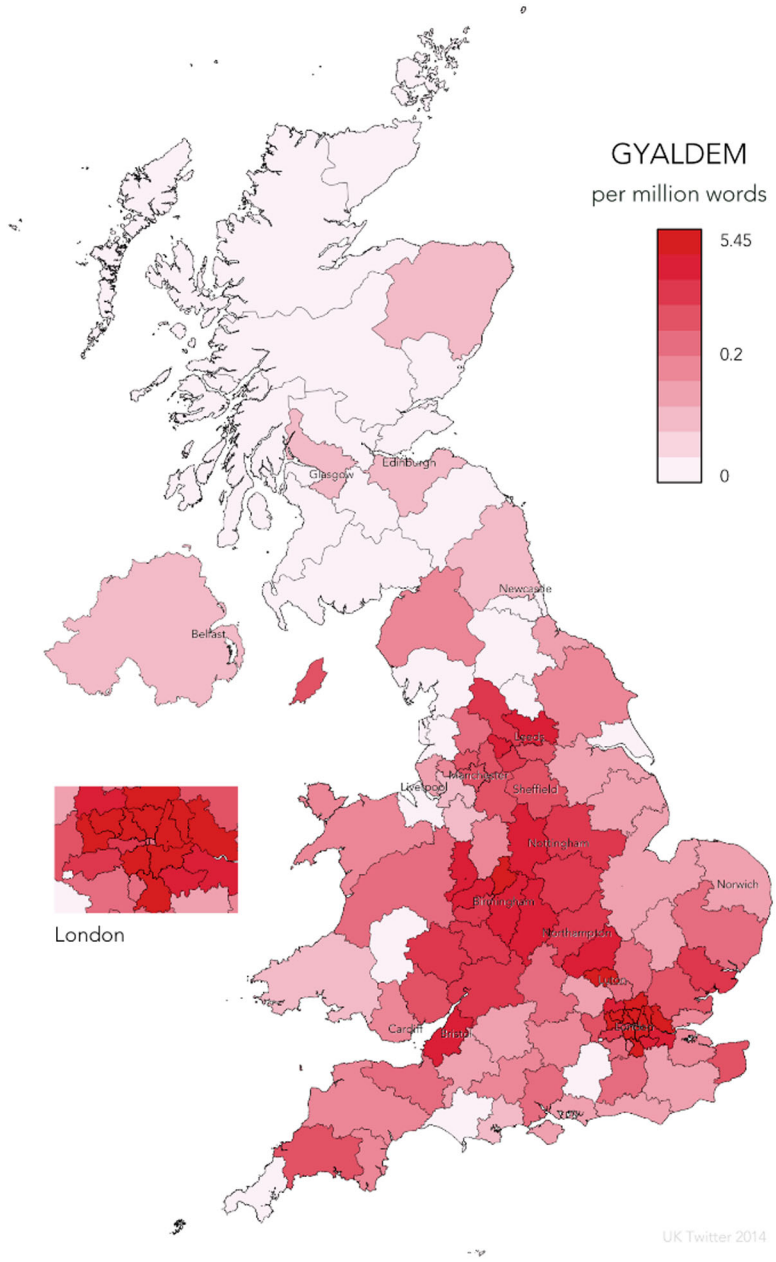

FIGURE 5 Frequency of 'gyaldem' across the United Kingdom per million words.

regional pattern in the set of 47 core MLE words. In other words, most of these 47 words follow the same basic regional pattern. We find that there is no clear secondary regional pattern or source of words associated with MLE or arguably with MBE more generally. Specifically, the first dimension accounts for almost half the variance in the dataset—just under 50% (.485)—whereas subsequent dimensions account for at most 5% of the variance in the dataset, as can be seen in Figure 6.

To visualise this regional pattern, scores for dimension 1 across the 124 postcode areas are shown in Figure 7. This map is effectively a hotspot of MLE Lexis in the corpus. Areas in red are those with

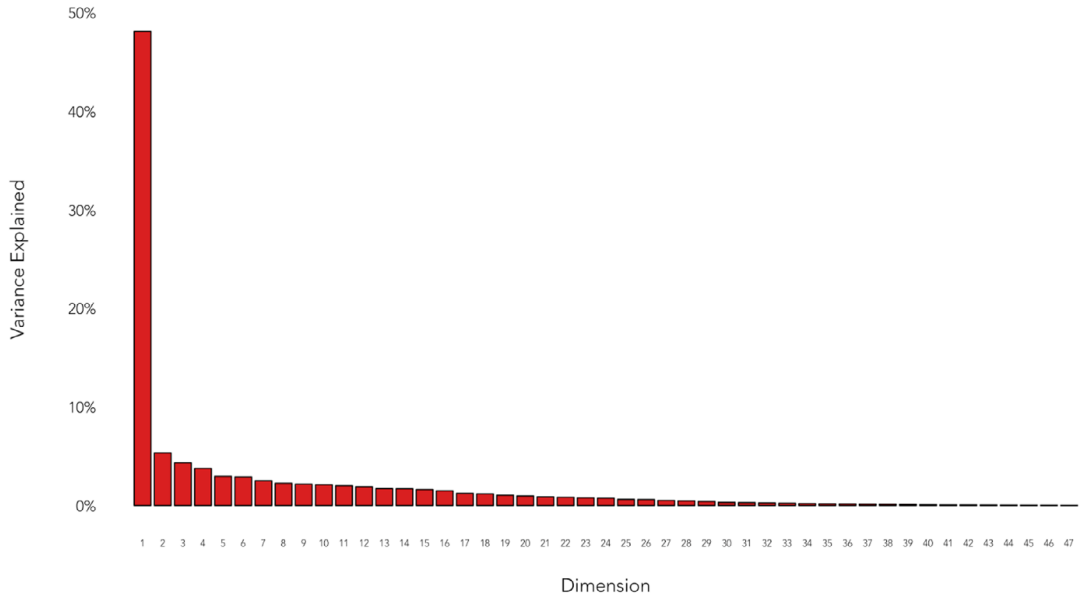

FIGURE 6 Amount of variance explained across principal component analysis (PCA) dimensions.

relatively high levels of MLE lexis, whereas areas in blue are those where these words are used less frequently. These results are clearly consistent with theories that place the origins of MLE in London, as this is where these dimension scores are strongest. In particular, the top four MLE lexical hotspots that we identify are the Greater London postcodes of IG (Ilford), EN (Enfield), N (North London) and CR (Croydon). In fact, 9/10 are London postcodes. The only location outside London in the top 10 is Luton. All these locations are extremely ethnically diverse and are home to relatively high numbers of people who identify as 'Black British' and 'Asian British'. We likewise find that the areas least strongly associated with MLE lexis are mainly rural areas that are geographically and culturally disparate from those where MLE lexis is most frequent. The bottom four regions—ZE (Shetland), ML (Motherwell), KW (Kirkwall) and HS (Outer Hebrides)—are in Scotland and are considerably less ethnically diverse than London and the South East of England. For instance, the Outer Hebrides is overwhelmingly 'White Scottish/British' (97.4%), with very few residents identifying as 'African/Caribbean Black' (0.07%) or 'Asian' (0.05%).

Our analyses also permit a deeper investigation of the origins of MLE in London. Although MLE has been documented in other areas of the city such as West London (e.g. Oxbury, 2021; Oxbury & McCarthy, 2019), there is evidence in our dataset that MLE is more strongly associated with neighbourhoods in the North and East of London. As the inset in Figure 7 shows, we do not see MLE lexis equally distributed across the city. The Western Central (WC) and South West London (e.g. SW) postcode regions are not as strongly associated with the MLE pattern when compared to East, North and (to some degree) South London postcode regions. This finding seems to support the claim that MLE is used mainly by individuals living and working in ethnically diverse, working-class inner city neighbourhoods (Cheshire et al., 2008, 2011; Fox, 2015; Fox et al., 2011). The WC region is a high-density commercial and administrative district with few permanent residents, whereas the SW postcode includes areas which are less ethnically diverse than other parts of London and is home to some of the most affluent neighbourhoods in the city. Thus, although MLE is concentrated in London, it appears to

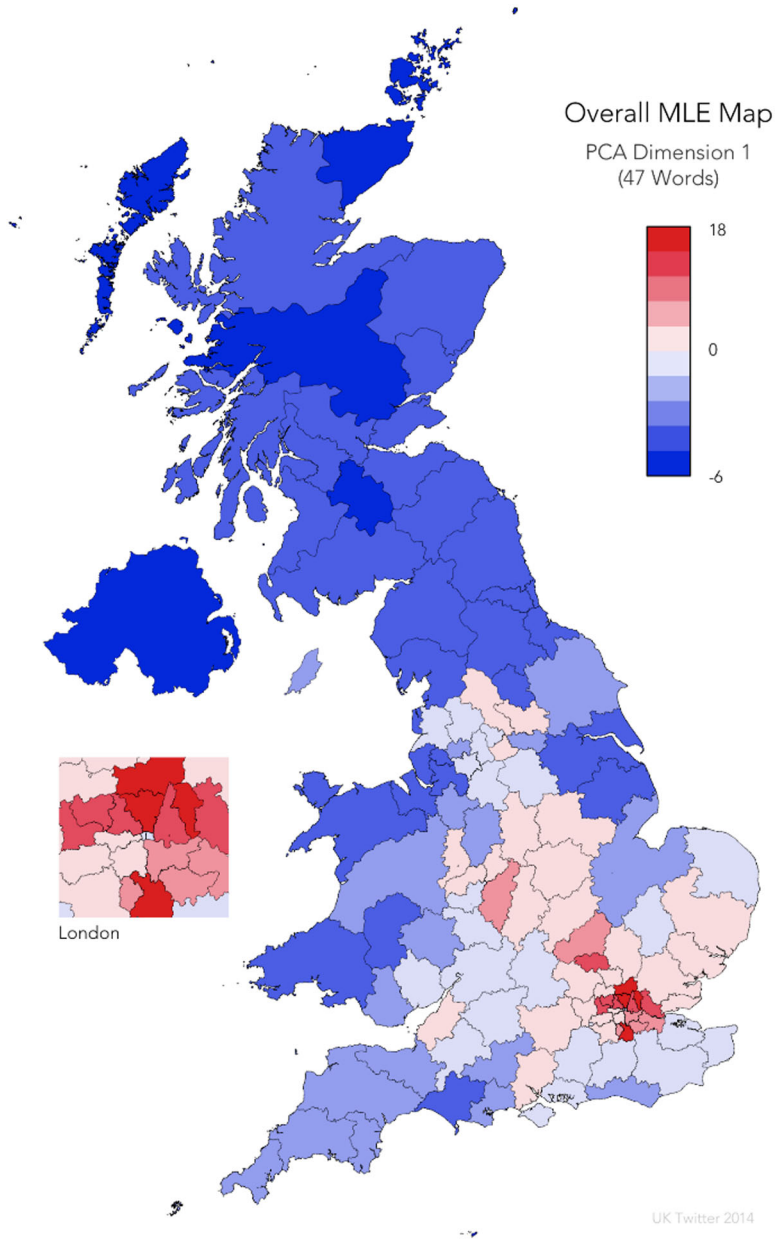

FIGURE 7 Visualisation of the principal component analysis (PCA) of Multicultural London English (MLE) lexis across the United Kingdom.

have originated in East and North London in neighbourhoods that are inner city, lower-SES and have higher numbers of people identifying as 'Black British' and 'British Asian'.

Our results also point to a relatively consistent area of spread of MLE lexis from London into other parts of England, especially in the general geographical vicinity of London and the South East of England, as well as other cities in the South and the Midlands. Following the four postcode areas in London, the next area that is most strongly associated with MLE lexis is Luton (LU), a large town some 30 miles from London. The other postcode areas in the top 20 outside Greater London are Milton

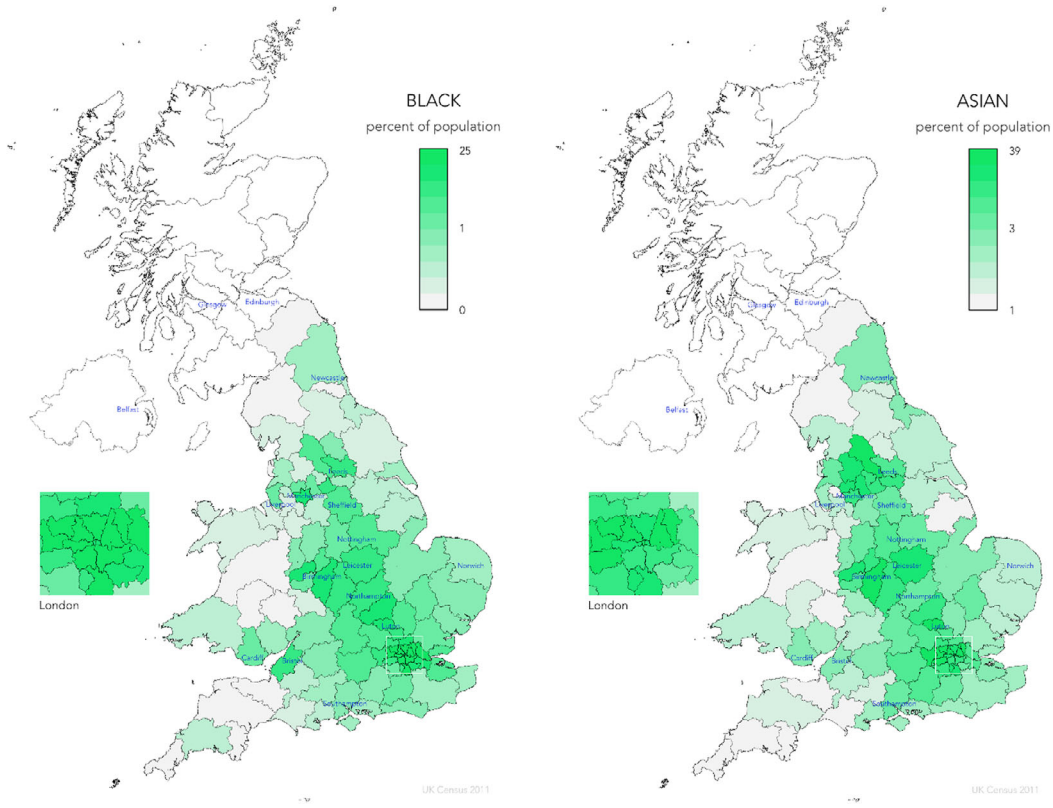

FIGURE 8 Per cent of the population in England and Wales who identify as 'Black' (left) or 'Asian' (right) according to the 2011 census.

Keynes, Birmingham, Northampton and Slough. All these are cities or towns that are relatively close to London, have diverse populations and are home to sizeable numbers of people identifying as 'Black British' and 'British Asian'. The most distant city in the top 20 is Birmingham, the second largest city in the United Kingdom and, like London, is characterised by very high levels of ethnic diversity. We therefore find that MLE lexis has predominantly spread northwards from London to nearby cities, up along the M1 and M6 motorways. Larger cities in the South of England, which are outside this M1 core, like Bristol and Southampton, are also characterised by somewhat frequent rates of MLE lexis. However, we do not see much spread of MLE outside of the South and Midlands, including into major cities like Manchester and Liverpool in the North of England, aside from the area around Leeds, Bradford, and Huddersfield. In fact, the linguistic patterns we identify here very closely mirror the ethnic composition of the United Kingdom. Figure 8 visualises the percentage of the population in England and Wales who identify as 'Black British' and 'British Asian' according to the 2011 census. As one can see, the figure—particularly the graphic that depicts the percent of individuals identifying as Black British—very closely resembles the distribution of the 47 core MLE words in Figure 7.

Perhaps surprisingly, however, we do not find evidence for MLE lexis in some areas where researchers have documented the use of MLE/MBE. Most notably, we identify a negative correlation between the MLE pattern and the North West English city of Manchester (cf. Drummond, 2018). One possible interpretation is that our data capture the earlier spread of MLE lexis, such that these words may not have yet been common in Manchester in 2014. In fact, there is some indication from Drummond's (2018) research that this may be the case. Although his respondents had similar vowel

TABLE 3 The table shows the correlation between Multicultural London English (MLE) words and Principle Component 1.

Rank	Word	PC1	Rank	Word	PC1
1	Ting	0.206335	25	paigon	0.148661
2	Fam	0.200315	26	sket	0.146314
3	Mandem	0.199381	27	booky	0.140337
4	Roadman	0.196076	28	duppy	0.13119
5	Clapped	0.195316	29	dench	0.128411
6	Bare	0.195169	30	hench	0.127753
7	Wasteman	0.194721	31	tekkers	0.107611
8	Gassed	0.192719	32	leng	0.106233
9	Yute	0.190964	33	beenie	0.098019
10	Rah	0.190126	34	rude boi	0.092816
11	Batty	0.186416	35	chirpse	0.091259
12	Endz	0.186272	36	prang	0.08973
13	Blud	0.182745	37	gully	0.077591
14	Yout	0.180465	38	yat	0.076611
15	Gyaldem	0.179493	39	nang	0.075995
16	Gyal	0.177507	40	cotch	0.071598
17	Alie	0.170981	41	uck	0.065756
18	Butters	0.169398	42	rudeboy	0.053417
19	Bredrin	0.166026	43	shotta	0.044326
20	Beef	0.166022	44	famalam	0.043842
21	Nitty	0.160263	45	slew	0.02705
22	Breh	0.156349	46	bora	−0.00506
23	Bruk	0.155341	47	ching	−0.00636
24	Yard	0.152937			

Note: Higher figures indicate a stronger correlation with the MLE pattern.

inventories and used similar consonantal features to those in the MLE studies, their awareness and use of lexical features appears to be more limited. For instance, Drummond's participants were unfamiliar with the MLE term *roadman* and many used dialect lexis specific to Manchester, such as *macca* 'rubbish or shit' (Drummond, 2018: 219). It is therefore possible that our data capture the actuation of the change before those items were adopted by young people in Manchester. This finding seems to suggest that the variety spoken in Manchester did not originate there, nor did it develop in parallel with MLE but rather it appears to have diffused from London.

Beyond the geographical spread of MLE, we can also use the PCA to identify words that are highly correlated with the MLE pattern (see Table 3). We find *ting, fam, mandem, roadman* and *clapped* to be strongly associated with the first component. This is perhaps expected given the status of these words in the variety. The word *ting* is an orthographic representation of TH-stopping (i.e. [t] for /*θ*/)—a feature found in both MLE and JE—which, in MLE/MBE, is lexicalised and is mainly constrained to the words *ting* 'thing', *yout* 'youth', and *teef* 'thief/steal' (Drummond, 2018; Ilbury, 2019). *Mandem* and *fam* are address terms and common in interactions with MLE speakers, and *roadman* is an enregistered racialised identity label, that is frequent in social media discourse (Ilbury, 2023). These words are also commonly used in grime music and other cultural practices associated with MLE (e.g. Adams, 2018).

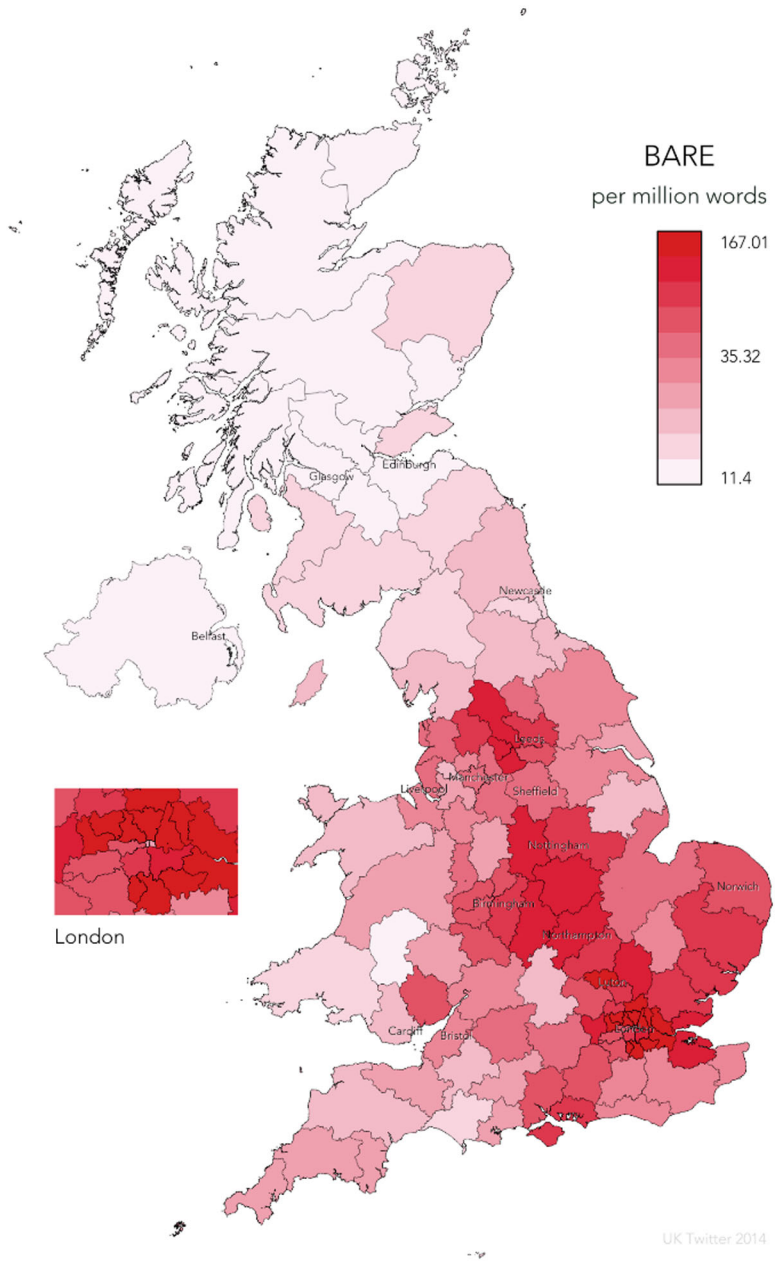

FIGURE 9 Frequency of 'bare' across the United Kingdom per million words.

Words which are not associated with the MLE pattern include *shotta, famalam, slew, bora* and *ching*. Most of these words are more obscure and/or dated or are associated more with other varieties. This is the case for *shotta* 'an independent man/drug dealer' which is more often considered part of JE.

Finally, we briefly return to the issue of polysemy. As we discussed in Section 3, many MLE words have SE equivalents. We noted there that it was not possible to distinguish between word senses in a dataset of this size. A possible issue then is that polysemy could skew the data. However, we do not find this to be the case. Figure 9 visualises the geographical distribution of *bare* across the United Kingdom.

In MLE, *bare* functions as an adjectival intensifier meaning 'lots of' as in 'there were bare people at the party'. Though this word is technically polysemous, we predicted that the SE sense, 'not clothed or uncovered', would be infrequent in the informal context of Twitter. Our suspicion is supported by the PCA: We find that *bare* is the sixth most linked word to the MLE pattern. In other words, it seems that it is primarily the MLE sense—not the standardised sense—that is being used on Twitter, at least with the strong underlying regional pattern. Indeed, when we map *bare*, we see the distribution of this word is comparable to a words that are not polysemous (e.g. *roadman*). This finding suggests that it is worthwhile considering the distribution of individual tokens rather than automatically discounting polysemous words (see also Grieve et al., 2019).

5 | DISCUSSION AND CONCLUSION

The previous analysis tracks the diffusion of a large number of lexical variants to infer the more general diffusion of an urban contact dialect across the United Kingdom. First focussing on individual words, then aggregated maps, we have identified and described the geographical trajectory of the diffusion of MLE. We find that

1. MLE lexis can be distinguished from heritage language influences (e.g. JE) and more general youth styles on the basis of its geographical distribution.
2. MLE appears to have first emerged in London, notably in the North and East of the city, where it was used most frequently by individuals in ethnically diverse, inner city neighbourhoods (see also Cheshire et al., 2008, 2011).
3. MLE appears to have diffused into other ethnically diverse urban areas in England (e.g. Luton and Birmingham) that are both geographically proximate to London and in which there are sizeable numbers of people identifying as 'Black British' and 'Asian British'. These areas form a geographical continuum from London providing a pathway for the spread of MLE.

These findings have important implications not just for our understanding of MLE but also for the study of urban contact dialects more generally. Assuming that theories of variant spread can be extended directly to the spread of varieties, we find that traditional models of linguistic diffusion cannot explain the patterns we observe. Notably, our findings contradict the assumption that linguistic innovations spread first from a large centre directly to another comparatively sized one, often skipping over smaller, geographically proximate areas (cf. Meyerhoff, 2022; Trudgill, 1974: Gravity model). Although we see evidence for the use of MLE lexis in Birmingham—the second most populous city in England (2014 pop. 1.09 million)—we find that words, such as *paigon, leng*, are more frequent in intermediate areas such as the town of Luton (2014 pop. 218,000). However, the area's relative proximity to London cannot account for the patterns we identify either (cf. wave model; Schmidt, 1872). Although there is some evidence for the progression of MLE lexis up through the M1/M6 motorway corridor, there are areas that are geographically disparate from London that show relatively high rates of MLE words (e.g. Leeds).

What then predicts the spread of this urban contact dialect? Our findings suggest that the spread of MLE is influenced *both* by population density and geography but also, and more importantly, the demographic similarity of regions. Notably, we observe higher rates of MLE lexis in areas that are urban, ethnically diverse, and home to greater numbers of lower SES households. Most notably, MLE lexis is used most frequently in areas where there are high numbers of people who identify as 'Black

British' and 'Asian British', to the point that the diffusion we identify largely mirrors the settlement pattern of the two communities in the United Kingdom.

This finding also helps explain why there are areas with relatively high numbers of people identifying as 'Black British' or 'British Asian' but with lower frequencies of MLE lexis. For instance, although 6% of Bristol (a city in the South West of England) identify as 'Black British', we see lower rates of MLE lexis here. Rather, it is JE that appears to be more frequently used in this area.

The low level of MLE lexis can be explained by fact that the city is somewhat geographically isolated from London and the areas surrounding Bristol are much less diverse than the city itself. Importantly, this observation not only illustrates the apparent linguistic diversity within a single ethnic category—in this case, 'Black British' (cf. Blake, 2014)—but it also supports our argument that it is *both* the ethnic and social composition of the area *and* its relative distance to London that predicts the spread of MLE.

Overall, we interpret our findings as support for the 'cultural model' of diffusion where words are seen to emerge in a culturally influential city and diffuse out into socially diverse areas (Eisenstein et al., 2014; Grieve et al., 2018). As in previous studies, which focus on American English, we find that although geographical proximity and population size may be important in the diffusion of linguistic innovations, the social and ethnic similarity of the city appear to play a more central role in the spread of MLE. We interpret these findings as evidence for the diffusion of MLE primarily through the social networks of Black and Asian users in London and their engagements with individuals in geographically proximate centres, and also their participation in shared cultural practices (e.g. grime; Drummond, 2018). Notably, however, the patterns that we identify do not provide any clear evidence for the lexical diffusion of MLE through social media, given that MLE appears to have spread incrementally via a geographical continuum from London. Although we acknowledge that (social) media is now likely a factor in the diffusion of MLE, it would appear that this is a more recent development (cf. Ilbury, 2023).

The arguments we have proposed here have major implications for sociolinguistic analyses of MLE and MBE more generally. Although earlier research on MLE found that it is an 'ethnically neutral' variety (Cheshire et al., 2011: 157), the present paper finds evidence to the contrary: The diffusion of MLE is mainly propelled through the social networks of Black and Asian users. We suggest that this distribution is indicative of the raciolinguistic enregisterment of MLE and its current status—at least at the level of lexis—as a variety that is broadly associated with Black and non-White speakers (also Boakye, 2019; Hirsch, 2018). Thus, although speakers from different ethnic and racial backgrounds may use MLE in practice, the distributional pattern of MLE lexis identified here appears to more strongly resemble the common perception of MLE as 'sounding Black' (see Drummond, 2016). Ideologies of MLE and its macro-level distribution, however, are clearly not unrelated since, as Rosa and Flores (2017) contend, varieties attain their coherence through ideologies that link linguistic features to particular types of people. What social personae are linked to MLE and how these relate to its diffusion remain questions for future research (though see Drummond, 2018; Ilbury, 2023; Walcott, 2022).

Additionally, we find little support for claims of a Multicultural *British* English (cf. Drummond, 2018). At least in 2014, MLE does not appear to be evenly distributed across the United Kingdom. Although MLE is frequent in parts of England, there are lower rates of MLE lexis in Northern Ireland and most of Scotland and Wales. Evidently then, MLE appears to be an *English*—not British—variety of English. This is most evident in areas which are ethnically diverse but are geographically disparate from London and outside of England, such as Glasgow, the largest city in Scotland some 400 miles from London. Although it is somewhat comparable to other areas where we see higher rates of MLE, it is urban and relatively ethnically diverse in comparison to the rest of Scotland (8% of residents identify as 'Asian', 2.4% identify as 'Black'), we do not see high frequencies of MLE lexis here. Clearly then, the spread of MLE lexis cannot be accounted for by the ethnic diversity of an area alone. Rather we

argue that this provides additional support for our claim that the diffusion of MLE is likely to have been propelled through the social networks of Londoners and their interactions with close-knit members of this network living in geographically proximate cities.[3]

Beyond our focus on MLE, the present paper provides some methodological insights in using social media data to explore large-scale patterns of language variation and change. First, our analysis demonstrates the analytical potential of using social media data to track the spread and diffusion of lexis. In the current paper, we show that there is a coherent and enregistered MLE lexical repertoire that is used outside of London and in geographically disparate areas where this variety has not yet been documented (e.g. Leeds). Second, our analysis demonstrates that some of the limitations of social media data (e.g. polysemy) may not be as problematic as initially perceived.

Nevertheless, we acknowledge that there remain several limitations to this approach. Notably, there is considerable orthographic variation in the representation of MLE words. For instance, the term *boukie* 'strange' is variably represented a <boukie>, <bouky> and <bookie>. A potential issue is that certain spellings may be more common in some areas than others. Focussing on a single spelling may then lead to a geographical bias for a particular form. This can be accounted for by mapping all orthographic variants of a word. When we do this for 'boukie', importantly we see the same pattern: All spellings are highly frequent in London and are correlated with the MLE pattern.

A final limitation that we mention briefly is that our analysis does not account for stylisations of MLE. As in other CS analyses, our approach assumes that tweets containing MLE words were produced by MLE speakers (e.g. Eisenstein, 2015; Grieve et al., 2018, 2019). However, research has shown that other speakers often stylise features of MLE—a practice especially common on social media (e.g. Ilbury, 2023). Although we acknowledge this as a limitation, given the size of our dataset and the consistency of the patterns we identify, it seems that such practices do not influence the data to a great extent, at least in 2014.

6 | CONCLUSION

In this paper, we have illustrated the analytical potential of employing a big data approach in exploring the diffusion and development of an urban contact dialect in tweets from millions of users across an entire nation. By identifying the social mechanisms that propel the diffusion of MLE, we have been able to identify the geographical trajectory of the spread and also explain how this variety diffused elsewhere. Our findings not only illustrate the central role of Black and Asian communities in the diffusion of MLE and in language change in the United Kingdom more generally but also call into question claims regarding the status of MLE as an ethnically neutral variety and its use across the United Kingdom as a whole (cf. Cheshire et al., 2011; Drummond, 2018). Given the implications of these findings, we join other scholars in calling for more research that employs big data in combination with other traditional sociolinguistic methods to better understand the status and development of contemporary patterns of language variation and change.

ACKNOWLEDGMENTS

The authors are extremely grateful to the editors, Sari Pietikäinen and Marie Maegaard, and two anonymous reviewers, for their extremely constructive feedback which has improved this article immeasurably. We would also like to thank Diansheng Guo for collecting the data and Devyani Sharma

[3] It is very possible, if not expected, that national identity plays a role in limiting the spread of MLE outside England. This is a question for future research to address.

for putting us in touch with each other. The research reported in this article was funded by the Arts and Humanities Research Council (UK), the Economic and Social Research Council (UK), Jisc (UK) (Jisc grant reference number 3154), and the Institute of Museum and Library Services (US), as part of the Digging into Data Challenge (Round 3). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

ORCID

Christian Ilbury https://orcid.org/0000-0001-9289-271X

REFERENCES

Aasheim, S. C. (1997). 'Kebab-norsk'-fremmedspråklig påvirkning på ungdomsspråket i Oslo. In U. Kotsinas , A. Stenström, & A. Karlsson (Eds.), *Ungdomsspråk i Norden* (pp. 235–242). Institutionen för nordiska språk.

Adams, Z. (2018). "I Don't Know Why Man's Calling Me Family All of a Sudden": Address and Reference Terms in Grime Music. *Language & Communication*, *60*, 11–27.

Agha, A. (2007). *Language and social relations*. Cambridge University Press.

Blake, R. (2014). African American and Black as demographic codes. *Linguistics Compass*, *8*(11), 548–563. https://doi.org/10.1111/lnc3.12118

Boakye, J. (2019). *Black listed: Black British culture exposed*. Dialogue Books.

Cassidy, F. G., & Le Page, R. B. (2002). *Dictionary of Jamaican English*. University of the West Indies Press.

Cheshire, J. (2013). Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics*, *17*(5), 608–633. https://doi.org/10.1111/josl.12053

Cheshire, J., & Gardner-Chloros, P. (2018). Introduction: Multicultural youth vernaculars in Paris and urban France. *Journal of French Language Studies*, *28*(2), 161–164. https://doi.org/10.1017/S0959269518000182

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2008). Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. In U. Ammon & K. J. Mattheier (Eds.), *Sociolinguistica: International yearbook of European sociolinguistics* (Vol. 22, pp. 1–23). Max Niemeyer Verlag.

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*(2), 151–196. https://doi.org/10.1111/j.1467-9841.2011.00478.x

Cheshire, J., Adams, Z., & Hall, D. (m.s.) Social meaning and the actuation of language change. London: Queen Mary University of London.

Christensen, M. V. (2012). 8220, 8210: Sproglig variation blandt unge i multietniske områder i Aarhus [Unpublished PhD thesis]. Århus University.

Clyne, M. (2000). Lingua franca and ethnolects in Europe and beyond. *Sociolinguistica*, *14*, 83–89. https://doi.org/10.1515/9783110245196.83

Collins, H. (2014). The second coming of grime. *The Guardian*. https://www.theguardian.com/music/2014/mar/27/second-coming-of-grime-dizzee-rascal-wiley

Dirim, I. (2005). Zum Gebrauch turkischer Routinen bei Hamburger Jugendlichen nicht-turkischer Herkunft. In V. Hinnenkamp & K. Meng (Eds.), *Sprachgrenzen uberspringen. Sprachliche Hybriditat und polykulturelles Selbstverstandnis* (pp. 19–49). Narr.

Drummond, R. (2016). (Mis)interpreting urban youth language: White kids sounding black? *Journal of Youth Studies*, *20*(5), 640–660. https://doi.org/10.1080/13676261.2016.1260692

Drummond, R. (2018). *Researching urban youth language and identity*. Palgrave Macmillan.

Drummond, R. (2021). *Multicultural British English*. https://www.robdrummond.co.uk/multicultural-british-english/

Eisenstein, J. (2015). Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics*, *19*, 161–188. https://doi.org/10.1111/josl.12119

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLoS ONE*, *9*(11), e113114. https://doi.org/10.1371/journal.pone.0113114

Fox, S. (2015). *The new Cockney: New ethnicities and adolescent speech in the traditional East End of London*. Palgrave Macmillan.

Fox, S., Khan, A., & Torgersen, E. (2011). The emergence and diffusion of Multicultural English. In F. Kern & M. Selting (Eds.), *Ethnic styles of speaking in European metropolitan areas* (pp. 19–44). John Benjamins.

Gadet, F. (2022). France: Youth vernaculars in Paris and surroundings. In P. Kerswill & H. Wiese (Eds.), *Urban contact dialects and language change* (pp. 264–281). Routledge.

Gates, S. M. (2018). Language variation and ethnicity in a Multicultural East London Secondary School [Unpublished PhD thesis]. Queen Mary University of London.

Grieve, J., Nini, A., & Guo, D. (2017). Analyzing lexical emergence in American English online. *English Language & Linguistics*, *21*(1), 99–127.

Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. *Journal of English Linguistics*, *46*(4), 293–319. https://doi.org/10.1177/0075424218793191

Grieve, J., Montgomery, C., Nini, A., Murakami, A., Guo, & D. (2019). Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence*, *2*(11), 1–18.

Grieve, J., Speelman, D., & Geeraerts, D. (2011). A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, *23*(2), 193–221. https://doi.org/10.1017/S095439451100007X

Grondelaers, S., & Marzo, S. (2023, firstview). Why does the shtyle spread? Street prestige boosts the diffusion of urban vernacular features. Language in Society, 52(2): 295-320.

Gross, J., & Boyd, S. (2022). Sweden: Suburban Swedish. In P. Kerswill & H. Wiese (Eds.), *Urban contact dialects and language change* (pp. 246–263). Routledge.

Hackney. (2016). *A profile of Hackney, its people and place*. https://hackney.gov.uk/population

Hall, D. (2020). The impersonal gets personal: A new pronoun in Multicultural London English. *Natural Language and Linguistic Theory*, *38*(1), 117–150. https://doi.org/10.1007/s11049-019-09447-w

Hirsch, A. (2018). *Brit(ish): On race, identity and belonging*. Penguin Random House.

Huang, Y., Guo, D., Grieve, J., & Kasakoff, A. (2016). Understanding U.S. regional linguistic variation with Twitter data analysis. *Computers, Environment, and Urban Systems*, *59*, 244–255. https://doi.org/10.1016/j.compenvurbsys.2015.12.003

Ilbury, C. (2019). Beyond the offline: Social media and the social meaning of variation in East London [Unpublished PhD thesis]. Queen Mary University of London.

Ilbury, C. (2020). 'Sassy Queens': Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of Sociolinguistics*, *24*, 245–264. https://doi.org/10.1111/josl.12366

Ilbury, C. (2023, firstview). The Rec. ontextualisation of Multicultural London English: Stylising the 'roadman'. *Language in Society*, 1–25.

Ilbury, C., & Kerswill, P. (2024). How multiethnic is a multiethnolect?: Recontextualising Multicultural London English. In B. Svensen & R. Jonsson (Eds.), *Routledge handbook on language & youth culture* (pp. 362–376). Routledge.

Jaspers, J. (2008). Problematizing ethnolects: Naming linguistic practices in an Antwerp secondary school. *International Journal of Bilingualism*, *12*(1–2), 85–103. https://doi.org/10.1177/13670069080120010601

Kerswill, P. (2013). Identity, ethnicity and place: The construction of youth language in London. In P. Auer, M. Hilpert, A. Stukenbrock, & B. Szmrecsanyi (Eds.), *Space in language and linguistics: Geographical, interactional, and cognitive perspectives* (pp. 128–164). Walter de Gruyter.

Kerswill, P., & Wiese, H. (2022). *Urban contact dialects and language change*. Routledge.

Khan, A. (2006). A sociolinguistic study of Birmingham English: Language variation and change in a multiethnic British community [Unpublished PhD dissertation]. Lancaster University.

Kossmann, M. (2017). Is Dutch Straattaal a mixed multiethnolect? A Moroccan perspective. *Applied Linguistics Review*, *10*(3), 293–316. https://doi.org/10.1515/applirev-2017-0050

Kotsinas, U. (1988). Rinkebysvenska—en dialekt? [Rinkeby Swedish—a dialect?]. In P. Linell, V. Adelswärd, T. Nilsson, & P. A. Petersson (Eds.), *Svenskans beskrivning 16* (pp. 264–278). Universitetet i Linköping.

Meyerhoff, M. (2022). Baby steps in decolonising linguistics: Urban language research. In P. Kerswill & H. Wiese (Eds.), *Urban contact dialects and language change* (pp. 145–157). Routledge.

Nortier, J. (2008). Ethnolects? The emergence of new varieties among adolescents. *International Journal of Bilingualism*, *12*, 1–5. https://doi.org/10.1177/13670069080120010101

ONS. (2011). *UK census data. Office of National Statistics*. ONS. https://www.ons.gov.uk/census/2011census/2011ukcensuses

Oxbury, R. (2021). Multicultural London English in Ealing: Sociophonetic and discourse-pragmatic variation in the speech of children and adolescents [Unpublished PhD thesis]. Queen Mary University of London.

Oxbury, R., & McCarthy, K. (2019). Acquiring a multiethnolect: The production of diphthongs by children and adolescents in West London. In Proceedings of the 19th international congress of phonetic sciencesc (pp. 2208–2212). Phonetic Association: London.

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, *12*(1–2), 43–61. https://doi.org/10.1177/13670069080120010401

Quist, P. (2022). Denmark: Danish urban contact dialects. In P. Kerswill & H. Wiese (Eds.), *Urban contact dialects and language change* (pp. 186–205). Routledge.

Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society*, *46*(5), 621–647. https://doi.org/10.1017/S0047404517000562

Schmidt, J. (1872). *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Böhlau.

Sebba, M. (1993). *London Jamaican*. Routledge.

Şimşek, Y., & Wiese, H. (2022). Germany: Kiezdeutsch. In P. Kerswill , & H. Wiese (Eds.), *Urban contact dialects and language change* (pp. 300–322). Routledge.

Svendsen, B. A. (2022). Norway: Contemporary urban speech styles. In P. Kerswill & H. Wiese (Eds.), *Urban contact dialects and language change* (pp. 206–222). Routledge.

Svendsen, B. A., & Røyneland, U. (2008). Multiethnolectal facts and functions in Oslo, Norway. *International Journal of Bilingualism*, *12*(1&2), 63–83.

The Economist. (2021). Grime and UK drill are exporting multicultural London English. *The Economist*. https://www.economist.com/britain/2021/01/30/grime-and-uk-drill-are-exporting-multicultural-london-english

Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, *2*, 215–246. https://doi.org/10.1017/S0047404500004358

Walcott, R. (2022). A Tweet at the table: Black British identity expression on social media [Unpublished PhD thesis]. King's College London.

Wiese, H. (2009). Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua*, *119*(5), 782–806. https://doi.org/10.1016/j.lingua.2008.11.002

Wiese, H. (2022). Urban contact dialects. In S. Mufwene, & A. M. Escobar (Eds.), *Language contact—Volume 2: Multilingualism in population structure* (pp. 115–144). CUP.

How to cite this article: Ilbury, C., Grieve, J., & Hall, D. (2024). Using social media to infer the diffusion of an urban contact dialect: A case study of Multicultural London English. *Journal of Sociolinguistics*, 1–26. https://doi.org/10.1111/josl.12653