“Good to see the mandem from the endz doing their ting”

Multicultural London English in the tweets of Grime artists

Ilari Raiski
Master’s Thesis
Master’s Program in English Studies
Faculty of Arts
University of Helsinki
2021

Tiedekunta/Osasto – Fakultet/Sektion – Faculty	Laitos – Institution – Department
Humanistinen tiedekunta	Nykykielten laitos

Tekijä – Författare – Author

Ilari Raiski

Työn nimi – Arbetets titel – Title

"Good to see the mandem from the endz doing their ting" – Multicultural London English in the tweets of Grime artists

Oppiaine – Läroämne – Subject

Englannin kielen ja kirjallisuuden maisteriohjelma

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä– Sidoantal – Number of pages
Pro gradu -tutkielma	Toukokuu 2021	59 s. + liitteet

Tiivistelmä – Referat – Abstract

Tämä sosiolingvistinen tutkielma pyrkii selvittämään Lontoossa kehittyneen kielivarieteetin, *Multicultural London Englishin* (MLE), käyttöä kirjoitetussa muodossa. Tutkielman tavoitteena on selvittää, käytetäänkö tämän niin kutsutun *multietnolektin* piirteitä myös kirjallisessa ilmaisussa. *Multietnolektit* sekoittavat monia eri kieliä ja yksinkertaistavat valtakieltä. *Multietnolekteja* on syntynyt moniin Euroopan monietnisiin kaupunkeihin vuosituhannen vaihteessa, ja ne ovat olleet sosiolingvistiikan tutkijoiden mielenkiinnon kohteita jo pitkään. Tämä tutkielma keskittyy kolmeen MLE:n piirteeseen: slangisanojen käyttöön, fonologisten piirteiden toteutumiseen kirjoitetussa muodossa sekä *man*-sanan käyttöön pronominina.

Aineistona tutkimuksessa käytettiin grime-artistien twiittejä. Twitter-aineistot ovat tulleet viime vuosina yhä suosituimmiksi kielentutkijoiden parissa. Tähän tutkimukseen se valikoitui, koska epästandardeja varieteetetteja esiintyy sosiaalisessa mediassa verrattain paljon. Aineisto kerättiin 20 grime-artistin twiiteistä, ja kokonaisuudessaan aineistoon kuuluu noin 60 000 twiittiä ja noin 600 000 tuhatta sanaa. Aineistoa käsiteltiin käyttäen *SketchEngine*-nimistä korpustyökalua. Löydöksiä verrattiin MLE korpukseen, joka koostuu puhutusta MLE-aineistosta. Suurin osa analyysista oli määrällistä, mutta kielenpiirteitä käsiteltiin esimerkkien avulla myös laadullisesti. Aineistosta etsittiin avainsanoja, sanojen yleisyyttä sekä niiden kollokaatiota (sanoja, jotka esiintyvät yhdessä).

Tutkimustulokset osoittavat, että monet MLE:n piirteet ovat siirtyneet puhutusta kielestä myös kirjallisen ilmaisun piiriin. MLE:n fonologisia piirteitä sisältäviä ilmaisuja löytyi aineistosta jonkin verran. Ilmaisut, kuten *dem, dat, da* ja *ting*, olivat yleisiä suurimmalla osalla artisteista, mutta monissa tapauksissa fonologisten ilmaisujen käyttö rajoittui muutamaan ihmiseen. MLE:n slangisanoja käytettiin aineistossa hyvin runsaasti, joten voidaan todeta, että niiden käyttö on vakiintunutta myös kirjallisessa muodossa. Myös *man*-pronominin käyttö oli hyvin yleistä tutkimusaineistossa. Sekä slangisanojen että *man*-pronominin käyttö oli yleisempää Twitter-ainestossa kuin puhutun kielen korpuksessa. Vertailu puhutun kielen korpuksen kanssa oli kuitenkin haastavaa, sillä se oli vanhempi kuin Twitter-korpus ja koostui nuorempien ihmisten haastatteluista. Tästä syystä jatkotutkimusta vaaditaan, jotta puhutun ja kirjallisen muodon välisiä eroja voidaan tutkia tarkemmin MLE-puhujien keskuudessa.

Avainsanat – Nyckelord – Keywords

Multicultural London English, multietnolektit, sosiolingvistiikka, Twitter, Grime, korpuslingvistiikka,

Säilytyspaikka – Förvaringställe – Where deposited

Helsingin yliopiston digitaalinen arkisto HELDA (E-thesis)

Table of Contents

List of Tables

1. Introduction

New language varieties have emerged in the turn of the 21st century in many European cities through language contact between various languages and varieties. In academia these varieties are often referred to as *multiethnolects*. Multiethnolects are varieties or styles that have emerged in multicultural areas, where people combine elements from a variety of languages. These urban slangs or varieties have been recorded widely across Europe in Scandinavia, Belgium, the Netherlands, Germany, and the United Kingdom (Cheshire et al. 2011, p. 152). *Multicultural London English* (MLE) is one of these multiethnolects and the one that is the topic of this paper.

MLE was born in the highly multicultural areas of East London through language contact between varieties of English and other languages (Kircher and Fox 2019b, p. 4). The main influencers of the variety are noted as being Asian, Caribbean, and African Englishes and local varieties such as Cockney (Kerswill 2014, p. 432). As there is a vast number of different languages that have affected and still affect MLE, speakers might sound quite different from one another. Therefore, the variety is best understood as an umbrella that holds a wide range of different linguistic features used in the multicultural areas of London (Cheshire at al. 2011, p. 154).

MLE has spread throughout London, the UK, and to some extent, even the world. Drummond (2016, p. 11) notes that MLE features have spread to at least Manchester and argues that the variety is spreading throughout the UK. On top of that, MLE features can now be heard worldwide thanks to the music genres, grime and UK drill, which have helped MLE features to spread around the globe (The Economist, 2021). The impacts of urban London and MLE can be seen in Finland as well. The so-called "roadman-style" has recently been popular among teenagers thanks to grime and Netflix's TV series, *Top Boy*, and nowadays, it is not uncommon to hear MLE slang terms/elements in the speech of Finnish teenagers. (Manner & Teittinen 2021).

As noted in the previous paragraph, the musical style, grime, has been mentioned widely as a motor for the spread of the variety. The language of grime is noted as being MLE. Grime MCs are avid users of MLE elements, and they even contribute new elements/expressions to the variety as well (Drummond 2018, p. 173). Hip-hop and rap have been widely used to study sociolinguistic aspects of African American English, so grime could be used similarly to study MLE. Adams (2018, p. 12) notes that "the significance of Grime for sociolinguistic research lies firstly in its close connection to British multi-ethnic dialects". Thus, this study uses the language of grime as a window for the features and development of the multiethnolect.

Many aspects of MLE have been studied during the last decades, such as the features of the multiethnolect (e.g. Cheshire et al. 2011), attitudes towards MLE (e.g. Kircher and Fox 2019b), and

the formation of the variety (e.g. Cheshire 2020). However, there has not been any research on MLE in written form. Social media has changed how written language is used; the language of social media is much closer to everyday language than it has ever been in written forms, thus offering a viewpoint to regional and social variation in language use (Nguyen 2019, p. 1). Features from many non-standard varieties of languages have made their way into the language of online communication (Jørgensen et al. 2015, p. 9), and this study tries to identify whether MLE elements are used online as well.

Research in online communications and social media especially has gained popularity in recent times. This study uses the microblogging website Twitter as the data source. The use of Twitter offers many benefits for linguistic research; the data is relevant as it is what people are actually saying and it can be collected in real time (Zappavigna 2011, p. 789). This study uses the tweets of 20 London-based grime artists as the material for analysis. It was difficult to tie down MLE speakers by using hashtags etc. Grime artists are noted as MLE users, so it seemed logical to select them as the subjects for this study. Furthermore, public figures' privacy is narrower; thus the use of their tweets is more sensible from an ethical standpoint as well. This study concentrates on three aspects of MLE: the use phonological elements, slang words, and the new pronoun innovation *man*. The research questions for this study are:

1. Are MLE elements used in written form?

2. Which features have made their way into written language?

3. Does the usage of elements differ between spoken and written mediums?

Section 2 provides the background for the research, explains key terms relevant to the study, and showcases previous research regarding MLE, multiethnolects, and other related topics. Section 3 presents the materials as well as the methods used to process and analyze the data. Section 4 introduces the results from the study along with an analysis of the findings. Section 5 discusses the findings in more depth, and section 6 concludes the thesis with an overview of the findings, limitations of this study, and suggestions for future research.

2. Background

The background section defines and discusses key terms used in the study and shows the findings of previous researchers in the research areas of multiethnolects, MLE and related fields. Section 2.1 gives an overview of multiethnolects, which is the term that many researchers have used to describe MLE, and similar developments in other countries. Section 2.2 discusses the history of MLE and some fundamental distinctions regarding the term *Multicultural London English*. 2.3 shows the attitudes towards MLE and section 2.4 maps out the features of MLE that previous researchers have found. Section 2.5 examines the links between the music genre, grime, and MLE, as that is of importance due to the data selection of this paper. Finally, section 2.6 inspects the language of social media as well as previous research on non-standard varieties in written form.

2.1 Multiethnolects

As mentioned in section 1, MLE is described as a multiethnolect of English by many researchers, and that is going to be the term used in this study as well. Michael Clyne coined the term *multiethnolect* in the year 2000 (Quist 2008, p. 44). According to Cheshire et al. (2015, p. 2), the term multiethnolect "refers to the way that in mixed multicultural neighbourhoods, young people may combine elements from different heritage languages with the dominant mainstream language". Quist (2008, p. 44) defines multiethnolects as a "linguistic ´something´, a variety or style, which has developed in multiethnic urban communities and which is associated with speakers of mixed ethnic groups." The fact that Quist describes multiethnolects, as a linguistic "something" says a great deal about multiethnolects – the term really cannot be described in the way that more traditional language varieties are described as. Obviously, the term is closely linked to terms such as sociolect (variety belonging to a particular social group[Macmillan Dictionary]).) and dialect (variety belonging to a specific area or community[Macmillan Dictionary]). However, multiethnolects are more heterogeneous than those aforementioned terms as the speakers of a multiethnolect cannot be described as precisely as one could with sociolects and dialects. It should also be noted that the term *ethnolect* has been used in linguistics for a longer time. The term is used in cases where there is only one language mixed with the mainstream language, and therefore it is important to make a distinction between ethnolects and multiethnolects (Quist 2008, p. 49).

Multiethnolects draw linguistic innovations from several minority languages/varieties and mix elements from those with the majority language. Therefore, multiethnolects are not a uniform entity but rather a mosaic of languages that have come into contact with the majority language. These varieties are constantly interacting together, making multiethnolects dynamic and shifting varieties. Multiethnolects have been recorded in many European countries e.g. The Netherlands,

Denmark, Sweden, Germany, etc. (Cheshire et al., 2015, p. 3-4). They have all emerged around the turn of the 21st century, and they have arisen from similar environments, highly multicultural urban areas.

The term is not used by all researchers who have worked with these new language varieties. Cheshire et al. note that some researchers opt to use terms such as language practices or repertoires (2011, p. 152). Others simply call these varieties *youth language* or *contemporary urban vernaculars* (Rampton 2015, p. 25). Even with the differences in terms of naming these new varieties, it is still evident that this way in which people (especially adolescents) draw elements from a pool of many other languages and varieties, needs to be addressed within a specific term. One could use other terms besides *multiethnolect*, but as it is well accepted among many researchers, it will be used in this paper as well.

The term multiethnolect can be problematic, so one must be cautious when using the moniker. Many researchers have used the term in differing ways, and therefore it is essential to note some distinctions about the concept. One key issue with the concept, as Quist (2008, p. 49) notes, is the matter of whether multiethnolects should be approached as varieties or stylistic practices. When a multiethnolect is considered a variety, it is contrasted to the host language. According to this idea, multiethnolects are forms of the "standard" national language. If multiethnolects are described as stylistic practices, the focus is more on identity and social meaning construction. Quist (2008, p. 50-51) describes stylistic practice as follows: "stylistic practice is the process through which signs and differences become meaningful resources in daily enterprises and activities. Stylistic practice covers the processes that connect different resources (linguistic and nonlinguistic ones) in meaningful relationships in association with the participants' identity negotiations".

For the purposes of this study, either viewpoint could be used, but viewing them as varieties seems more logical. Quist (2008, p. 49) notes that: "The two approaches thus answer different kinds of questions. The variety approach contributes to the description of variation in the speech community, and it helps us to find out if there is anything linguistically systematic going on at all. The stylistic practice approach, on the other hand, is a way to examine the social meanings, functions, and consequences of the speech of the adolescents". The variational approach is closer to the purpose of this study, which is to find out systematic features of MLE in written context. Therefore, in this study, multiethnolects are perceived and referred to as varieties of a language.

It should be noted that the use of a multiethnolect (like in the case of MLE) can differ from speaker to speaker, but there are still fundamental linguistic features that set multiethnolects apart from other varieties of a language. Cheshire et al. (2011, p. 151) note that the speakers of a multiethnolect do not mix features from various languages with the heritage language, per se, but

rather the speakers have a wide "feature pool" from where they draw their linguistic innovations. Cheshire et al. also note in the article that "both dialect contact and language contact situations as producing a 'feature pool' from the range of input varieties, with speakers selecting different combinations of features from the pool, sometimes modifying them into new structures in the output varieties" (2011, p. 176). One key feature regarding multiethnolects is the fact that they are used to describe a culturally neutral language variety (even though, oftentimes they are not perceived as such) that has core elements in terms of phonetics, grammar, and discourse pragmatic markers (Cheshire et al. 2015, p. 3). Clyne (2000, p. 87) also notes that "several minority groups use it collectively to express their minority status and/or as a reaction to that status to upgrade it". As well as that, multiethnolects are spoken by non-immigrant monolinguals as well as people from bilingual backgrounds (Cheshire et al., 2015). This phenomenon, where members of the dominant group shift to use the minority variety, is a new way to express a special type of group identity (Clyne, 2000, p. 87, p. 2).

The definition of a multiethnolect can seem ambiguous. However, so is the phenomenon, and *multiethnolect* is an appropriate way to describe how, most often adolescents, combine elements from a large pool of different languages and varieties available to them. "Multiethnolects often emerge in the informal spontaneous interaction of linguistically diverse multiethnic peer groups and are therefore best considered as characteristic of particular communities of practice rather than of a bounded speech community" (Cheshire 2020, p. 310). Speakers of a multiethnolect cannot be categorized as one homogenous group, nor can the way in which they use the language, as they draw innovations from a repertoire consisting of many languages that varies from speaker to speaker.

2.2 Multicultural London English

2.2.1 History

The formation of a multiethnolect was explained in the previous section, and MLE, as a multiethnolect, obviously has formed as multiethnolects do. Especially in the eastern parts of London, there has been a considerable amount of immigration throughout the history of city. However, the rate of immigration has seen a drastic increase in recent decades (Cheshire 2020, p. 310). East London is seen as the birthplace of MLE by many researchers, as many different minorities have resided there, enabling features from many languages/varieties to interact together.

MLE has been most influenced by Caribbean creoles, African languages as well as Asian Englishes (Kerswill 2014, p. 432). There are many elements from vernacular English in MLE as well (Kerswill 2013a, p. 5). There is evidence of MLE characteristics in the 1980s (Kerswill 2014,

p. 435), but the variety really came to fruition at the turn of the century (Cheshire et al. 2015). Kerswill (2014, p. 435) claims that the first report of MLE was published in 1986 by Roger Hewitt. The article titled *White Talk Black Talk* dealt with the phenomenon where non-migrant Londoners started to borrow slang features, mostly from the Jamaican creole Patois. However, many of the features mentioned in the article are not regarded as features of MLE nowadays. Furthermore, the article describes the use of this vernacular as quite infrequent, and at that time, it was only used as an in-group variety. The linguistic elements used by adolescents in that time were still very close to the traditional Cockney dialect. Therefore, one could argue that people were influenced by features from foreign languages in the 1980s, but the formation of a definite linguistic variety had not been established yet.

Even though many immigrants were moving to London from the 1960s onwards, particularly in East London, the ethnic groups rarely interacted with one another (Cheshire et al. 2013, p. 66). In the 1980's, it was common that the ethnic groups used their own variety of English when interacting with their own ethno-social groups, and when conversing to others, they would shift to a more traditional version of London English. Sebba (1993) recorded this phenomenon, where Afro-Caribbean migrants would shift between patois and London English depending on whom they were talking to. The variety, which they spoke during that time, could be described as an ethnolect, but the process where a multiethnolect emerges, had not happened yet. Nowadays, Cockney, which has long been the most frequently used form of slang in London, has seen a radical decline in use (Cheshire et al. 2011, p. 164). Many researchers (e.g. Kerswill 2014; Cheshire et al. 2013) note that MLE is replacing Cockney and that the only speakers of Cockney are older people – adolescents have shifted to use MLE.

2.2.2 Definition of MLE

Section 2.1 showcased the problems using the term *multiethnolect* and how one defines that, and the case is very similar with the label, MLE. As mentioned, this study discusses multiethnolects, so MLE as well, through the variational approach. Cheshire et al. (2011, p. 154) use the term MLE "to refer to the overall range of distinctive language features used in multiethnic areas of London, conceptualising MLE as a repertoire of features." The idea of a "feature pool" with multiethnolects was mentioned in section 2.1, and this is how MLE is going to be regarded in this study. MLE is used to refer to London's urban vernacular speech style, taking into account the diverse and changing nature of the multiethnolect.

This paper uses MLE to describe this new variety, but it should be noted that one must be considerate when using the term (similarly with the use of multiethnolect). Drummond (2016)

introduces the term Multicultural British English (MUBE) as an extension of MLE, to consider the spread of the variety. In his study, Drummond (2016) found that MLE features had spread to speakers from Manchester. This finding supports, the idea that MLE is spreading, and thus a moniker such as MUBE would be more appropriate to describe the phenomenon. However, there has not been enough research to establish MUBE as a specific concept (Drummond 2016, p. 644). Therefore, MLE will be the term that will be used in this paper, as it is widely used and accepted, but this is not done without caution. It should also be understood that MLE is not a language variety that is only spoken in London, as MLE features have spread domestically and even internationally.

2.3 Attitudes towards MLE

Attitudes towards the variety have been one of the most researched areas of MLE. Even though this paper is not dealing with the attitudes towards MLE, it is still essential to understand the prevailing stances that people have towards the multiethnolect in order to understand perceptions that people often have about non-standard varieties. As well as that, MLE speakers' own attitudes about the multiethnolect are also presented to understand better the reasons why MLE speakers use the variety.

Multicultural London English is the academic name for this new multiethnolect, but in general discourse and mainstream media, the variety is often discussed with other monikers, such as *Jafaican* (Fake Jamaican) or *black street patois* (Drummond 2017, p. 2). Both of these terms have the perception, that the speakers would be ethnically black, even though, as noted, the speakers of MLE are not from just one ethnic group. This belief is harmful to the speakers of MLE because, as Lippi-Green (2012, p. 253) notes, "Varieties 'linked to skin that isn't white' tend to be particularly stigmatized". So, the terms, in which the variety is discussed, carry incorrect assumptions about the speakers of the multiethnolect. As well as that, overall opinions about the variety are mostly negative as well.

Non-standard varieties of English are generally looked down on, and so is MLE. Kircher and Fox (2019a) note the negative perceptions towards MLE in media representations and the fact that many laypeople see speaking MLE as a hurtful phenomenon. Drummond (2016, 641) examined comments from mainstream media and other commentators and found comments that say that multiethnolect speakers in the UK "are not speaking properly", "sound ridiculous" or "are literally talking their way into unemployment". As noted, people often associate MLE with black people. Moreover, Kircher and Fox also note that the variety is most often associated with young, working-class males from East London. In truth, the speakers of MLE can be found in all socio-economical classes, age groups, and boroughs of London (2019b, p. 4).

Standard language ideology (SLI) is a fundamental concept when discussing attitudes regarding minority varieties. Standard language ideology is a belief of an identifiable and stable language variety "that is inherently correct and, relatedly, leads to better communication among the masses" (Davila 2016, p. 128) Milroy (2001, p. 535) notes that most people live in societies that support this ideology, where varieties are classified as standard or non-standard. The ideology is deep-rooted in our society; the education system, media, and the corporate sector are all factors that support the idea where the imaginary concept of a *standard language* is seen as the correct way to speak, and other varieties are seen as incorrect.

This belief is not only wrong, but also wounding to the speakers who do not buy into this ideology. Speakers of non-standard varieties face difficulties in education, employment, the judicial system as well as when they are searching for housing – speaking a multiethnolect, or any other non-standard variety, puts one at a disadvantage in various everyday situations (Kircher & Fox 2019a, p. 580). Another study by Kircher & Fox (2019b) studied SLI and attitudes towards MLE in London. They found out that SLI is far more pervasive with non-MLE speakers. Non-MLE speakers described MLE as *broken*, *incorrect* etc., demonstrating the fact that they feel that MLE is improper in comparison to *standard language*. The same study also noted that MLE speakers recognized the mixed nature of MLE speakers, whereas non-MLE speakers had far more stereotypes towards MLE speakers.

As noted in section 2.1, the speakers of a multiethnolect use the variety to express their belonging to a particular group. Drummond (2017, p. 657) notes that speakers use MLE features to index masculinity and being young and being from the *street*. Most MLE speakers are very much able to switch between variants (e.g. from MLE to more Standard British English [SBE]) in different contexts (Kircher & Fox 2019:b, pp. 4-5). The reasoning behind MLE speakers' decision to use MLE elements in certain contexts and use more *Standard English* in others is certainly of interest. This phenomenon will be further discussed in section 5.

2.4 Features of MLE

As noted in section 2.1, the features of MLE cannot be described in an ordinary fashion, as the usage of the multiethnolect differs from speaker to speaker. However, there are some key features that have been found in previous research. This section showcases features that previous researchers have found to be part of MLE. These features are also compiled into lists, presented at the end of this section (Tables 1 and 2).

2.4.1 Phonology

MLE differs from other British English varieties in terms of phonology, and the changes in the vowel system have been noted by several researchers. Cheshire et al. (2011, p. 158) claim that the most notable changes that have happened within MLE are found within the vowel system. One of the most notable changes is *Goose-fronting*, which is a key feature of MLE speakers. *Goose-fronting*, as the name in itself suggests, means that the vowel /u:/ in e.g. the word *goose* is pronounced in the front of the mouth. The phenomenon of *goose-fronting* is a global phenomenon, and it has become a feature of MLE through language contact (Cheshire et al. 2011, p. 171). The fronting of /u:/ in not the only discrepancy – there are significant changes that have happened within MLE speakers in terms of vowels and diphthongs. Cheshire et al. (2011, p. 160) also note backing of /æ/ (realized as /a/), and backing of /ʌ/ (realized as /ɑ/ or /ʌ/). Kerswill (2014, p. 433) also notes the important changes in the diphthong system, where monophthongs or narrow diphthongs replace broad diphthongs of Cockney (e.g. in FACE and GOAT, /æɪ/ → /eɪ/ and /ʌʊ/ → /oʊ/, respectively). One very interesting finding in the study of Cheshire et al (2011, p. 171), is that multicultural Londoners seem to adapt phonological features at a very young age, even before the children have attended school. This might be down to the fact that their caregivers do not speak English at home.

Vowels have undergone significant changes, but there are differences with consonants as well. Kerswill (2014, p. 433) notes the backing of /k/ before low back vowels (realized as /q/) and the reversal of h-dropping. Examples of where /k/ is backed to a /q/ sound can be found in words like *car*, *talk* and *cousin* (Kerswill 2013b). H-dropping, a very common feature in Cockney, is a phenomenon where word-initial /h/ is not pronounced in words like *hair* or *hand*. However, MLE speakers have reversed this phenomenon, and tend to pronounce the /h/ (Cheshire et al. 2008, p. 16). In this way, MLE has actually moved closer to Standard English.

Another notable pronunciational feature is that MLE speakers often replace dental fricatives /ð/ and /Θ with other phonemes. This phenomenon can be dissected into three features, according to Kerswill et al. (2007, p. 6). Firstly, Th-fronting refers to the phenomenon where /Θ/ is replaced by /f/ (e.g. in words *thin* and *bath*). The second phenomenon is Dh-fronting, where /ð/ is changed for /v/ in non-initial positions (e.g. in the word “mother”). The third element is Dh-stopping, which is when word-initial /ð/ is pronounced as /d/ (e.g. them → dem, then → den). In a later study, Kerswill (2018, p. 173) notes a fourth development regarding dental fricatives, Th-stopping. In terms of Th-stopping, the /Θ/ sound is realized as /t/ (e.g. thing → ting).

2.4.2 Morphosyntactic and discourse features

Other than the phonological features, the study conducted by Cheshire et al. (2011) lists three distinctive characteristics of the MLE feature pool that have been borrowed from other languages or varieties of English:

1. The new quotative expressions. The new quotative is formed by the phrase *this is + the speaker* (e.g. This is me "I'm from London".) The feature was found solely in the data of adolescent Londoners, which signifies it as a feature of MLE. The second new quotative is a globally spread expression *be like* (e.g. I was like "No, that's not me".). The expression is a part of many varieties of English, as it is of MLE as well.

2. The use of *be* in past tense (*was/were*) has acquired special features within MLE speakers. Speakers of non-standard varieties of English often use the past tenses of *be* in a non-standard way, as MLE speakers do as well (e.g. You was, we was etc.). However, what is interesting, is that MLE speakers seem to have developed a special pattern of usage for the past tense of *be*. "MLE speakers tend to level the forms to *was* and *wasn't* throughout the paradigm, instead of the widespread levelling to *was* and *weren't*" (Kerswill 2014, p. 433).

3. The third feature is the changes in definite and indefinite articles. MLE speakers have also acquired a feature in which they have simplified the allomorphy of articles. Speakers use pre-consonant forms of articles (*a* [ə] and *the* [ðə]) in pre-vocalic positions as well. Traditionally, one would use *an* [ən] and *the* [ði] in pre-vocalic positions, but MLE speakers can use the pre-consonant forms in both contexts (e.g. *an apple → a apple*).

On top of the aforementioned features, found by Cheshire et al. (2011), there are other innovations in MLE as well. Kircher and Fox (2019a, p. 849) also mention the use of discourse markers *innit* and *like* as features of MLE. The use of *innit* (abbreviation of isn't it) has been recorded by Torgensen et al. (2011) and Martínez (2015). The studies found that young Londoners use *innit,* as an invariant tag-question or a follow-up tag. The use is very flexible; *innit* does not follow the rules of regular tag question formation. It is most often used in order to seek confirmation to their utterance or to keep the listener's attention during a conversation. *Innit* is a feature of Cockney and many other British varieties of English, but the use of *innit* among MLE speakers is more frequent, and they have also acquired a new use for the word. *Innit* is used to foreground information; "a speaker who utters 'the girl *innit*' makes it clear that she intends to tell a story about something the girl in question did" (Cheshire et al., p. 14).

MLE speakers have also developed a new pronoun, man (Cheshire et al., 2013 and Hall 2020). The word *man* can be used in place of all other pronouns (1st, 2nd and 3rd person singular, as

well as 1st, 2nd and 3rd person plural), and it can be used as an impersonal pronoun (Hall 2020, p. 117). The word, when used as a pronoun, has lost it meaning referring to *males*, and the way in which MLE speakers use the word is very wide-ranging. This is visible from examples from the MLE corpus: “I blocked too many man”, “They call up their guys yeah, bare man outside school blud”, “Aah man that’s long that’s kind of long” (Hall 2020, p.121).

The use of intensifiers is another development within the speakers of MLE. An intensifier is “any device that scales a quality, whether up or down or somewhere between the two” (Bolinger 1972, p. 17). So, intensifiers are used to convey a message in a more expressive way. MLE speakers use the more common intensifiers (e.g. *very*”, *pretty* and *really*), but the study conducted by Núñez-Pertejo and Palacios-Martínez (2018) proved that there are also new intensifiers in MLE; *bare* and *proper*. *Bare* as an intensifier means “lots of” or “very” (e.g. “Man has bare p’s” or “He’s bare good, innit?”). *Proper* as well as *bare* is used as a very general intensifier (e.g. “That is not a proper job”).

Negation is another area where MLE speakers differ from speakers of SBE. Martínez (2013, p. 211) names three features of negation that are specific to MLE speakers: *ain’t, never,* and *multiple negation*. *Ain’t* is a non-standard way to mark negation, used especially in American English, and it replaces clauses *am not*, *is not*, *are not*, has not and *have not* (Palacios 2013, p. 213). *Never* expresses universal temporal negation in many varieties of English nowadays. *Multiple negation* refers to the use of two or more negatives in the same clause. *Multiple negation* is also used widely across non-standard varieties of English (Palacios 2013, p. 222).

Changes in vocabulary, especially in terms of use of slang words, are one area where there are considerable differences when compared to standard English speakers (Kircher & Fox 2019a). The influence of Caribbean creoles, predominantly Jamaican, has been noted in several publications, but Green (2014, p. 70), claims that the US slang has had even more of an influence on MLE. Furthermore, local varieties such as Cockney and other forms of London slang have provided slang terms to MLE, with many words also developing independently to the lexicon of MLE. Address and reference terms (word, phrase, name, or title used to address someone) are one area where slang terms are very visible. (Adams 2018, p. 23) notes variations of *brother* (*brudda*, *bre*, etc.), *fam*, *blud*, and many others as address terms that are associated with MLE. Section 4.3 features a list of MLE slang words, but it should be noted that list is not comprehensive. See Nott’s (2013) MLE glossary for a more extensive record of MLE slang expressions.

2.4.3 Lists of features

Tables 1 and 2 present the features of MLE. The features are compiled from studies of previous research papers regarding MLE. Table 1 contains phonological features of MLE, and Table 2 presents the morphosyntactic and discourse features.

Table 1. Phonological features of MLE

FEATURE	**EXAMPLE**
Fronting of /uː/ → /ʉː/ or /ʉː/	*Goose*
Backing of /æ/ → /a/	*Trap*
Backing of /ʌ/ → /ɑ/ or /ʌ/	*Strut*
Diphthongs	/æɪ/ → /eɪ/ in the word *face*
Reversal of H-dropping	*Hammer* pronounced as *hammer* (not '*ammer*)
K-backing	/k/ → /q/ in *cousin*
Dh-stopping	/ð/ → /d/ in *them*
Th-stopping	/Θ/ → /t/ in *thing*
Dh-fronting	/ð/ → /v/ in *mother*
Th-fronting	/Θ/ → /f/ in *bath*

Table 2. Morphosyntactic and discourse features of MLE

FEATURE	**EXAMPLE**
Quotative expressions	*Be like, this is + speaker*
Past tense *be*	*You was*
Articles	*A apple*
Man as a pronoun	*Man knows dat*
Discourse markers	*Innit, like*
Intensifiers	*Bare, proper*
Negation	*Ain't, never, don't*
Slang words / expressions	*Ends, butters, peng*

2.5 Grime & MLE

It is important to note a few things about grime, as that is going to be relevant in terms of the data selection for this study. Grime is a musical style that emerged in the UK at the turn of the 21st century. The genre combines elements from hip-hop drum 'n' bass, dancehall, and jungle (Adams 2018, p. 442), and it has been noted as "the most significant musical development within the UK for decades", according to a study conducted by Ticketmaster (Holden 2017). The same study compared the influence of grime to that of punk in the 1970s. Grime is an imperative part of today's

youth culture in the UK, especially in the urban parts of London, as that is the birthplace of grime.

Grime and MLE are linked to one another quite closely. Both of them emerged in the same place at the same time. Adams (2018, p. 12) claims that "the socially exclusive, racially inclusive nature of MLE mirrors grime, which suggests that the genre may be an alternative way of understanding MLE and interpersonal relationships among adolescents". This notion is supported by a study conducted by Dedman (2011), which found out that participants did not associate grime with race, but rather social class; "Grime was perceived as belonging to those who have survived London's deprived neighborhoods" (Adams 2018, p. 12). Similar to MLE, grime is associated with young black males, but in truth, neither of the phenomena are racially exclusive. There are only a few studies regarding grime in sociolinguistics. However, there are numerous studies about the influence that African American Vernacular English in rap/hip-hop have had on adolescents worldwide (Adams 2018, p. 12). Adams argues that grime is as viable as rap/hip-hop for sociolinguistic research, as grime is heavily linked to UK's multiethnolects. Adams (2018, p. 13) offers a quote from poet Charlie Darker to support his claim:

> The sound of a young London stands as a cultural landmark showcasing the thoughts, aspirations and speech of a generation of young people lost between the cracks of the system, perhaps quick to rebel but lost without a cause. The machine gun flow of the Grime MCs who began littering the airwaves over a decade ago breathed new life into language with a style of slang that was distinctly Urban, edgy and British. (Adams 2018, p. 13)

The definition of MLE in the Urban Dictionary (2012) also states that MLE "Originated in London (due to be such a multicultural area) and **quickly spreading to other areas of the UK through use and also through grime music**". So, it could be argued that the rapid spread of MLE has been achieved, at least partly, thanks to grime. The influence of British rap music as a disseminating factor of MLE has been noted at least in Australia (The Economist 2021).

"The language used by grime MCs is the vernacular language of many young(er) Londoners in the twenty-first century" (Adams 2019, p. 442). Drummond (2018, p. 173) also notes that grime MCs likely feed new elements into MLE as well as reinforcing existing elements. Adams (2019, p. 438-439) states that grime showcases the sociolinguistic developments that have happened in London and that grime actually expresses the modern identity of being a Londoner, or furthermore English. Therefore, this study will use the language of grime as a viewpoint to study the use of MLE. One could argue that MLE and the language used in grime are not the same thing. Drummond (2018, p. 174) notes the problem of distinguishing some linguistic features as being part

of MLE or grime specific language use. However, Drummond notes that the language of grime is commonly seen as MLE and that given the right circumstances, the two phenomena can very well be studied together.

2.6 Non-standard varieties in written form

As mentioned, there is no research about the usage of MLE in written context. However, there is plenty of research regarding other non-standard varieties and their presence in social media contexts. Therefore, findings that previous researchers have made with other language varieties are showcased here to better understand the phenomenon of non-standard varieties in social media.

Features of dialects rarely end up in formal language; however dialects are widely used in social media (Jørgensen et al. 2015, p. 9). Even though this thesis does not discuss dialects, per se, we can still safely presume that multiethnolects, as non-standard varieties of a language, follow the same pattern as dialects. Jones (2015, p. 431) notes the informal nature of social media, more specifically in this case, Twitter, where users "tend to write as they speak". Therefore, social media platforms are an ideal place to look for informal written language. There are many advantages of using social media as a data source, as data can be collected in real time and social media offers a look into the relationship between users as well as the identity portrayed in one's social media posts (Seargeant & Tagg 2014, p. 5).

The use of African American Vernacular English (AAVE) in social media has been recorded in many studies (e.g. Jones 2015, Blodgett et al. 2016). The study of Jones (2015) showcases the widespread use of phonological elements from AAVE in Twitter and indicates that non-standard varieties have a strong foothold on social media platforms. In many ways, the situation with MLE in Britain is equivalent to that of AAVE in the USA. Both are frowned upon and seen as incorrect. Both are also heavily linked with black speakers, AAVE even more so than MLE. Therefore, the studies that have studied AAVE in social media can be helpful to use as a reference point regarding this study as well. AAVE elements, especially phonological ones, are used a great deal on Twitter by speakers of AAVE (Jones 2015, p. 431). Therefore, it is interesting to see whether MLE users have developed their own patterns to use their variety in online communication.

The use of social media platforms in linguistics has gained popularity recently due to its conveniences: social media offers us a look into the most recent changes in languages as the data can be collected in real-time and the language used in social media tends to be colloquial in nature (Seargeant and Tagg 2014, p. 14). Social media users are a wide-ranging group, but the general users are young, the group in which language changes are most visible (Page et al. 2014, p. 18).

MLE is also a variant spoken mainly by young people; this is why a Twitter dataset will be a fitting way to seek answers to the research questions.

3. Materials and methods

This section presents the data as well as the methods that were used. The way in which this study will try to answer the research questions is through corpora. As the aim of this study is to identify MLE elements in written context, there will be a dataset of written language which is compared to a spoken language corpus. This section presents the data and the methods that were used to analyze the data.

3.1 Data

The spoken language corpus used in this study is the Multicultural London English corpus. It was compiled by Jenny Cheshire et al. in 2011 (MLE Corpus 2011). The corpus consists of interviews of 1 or 2 people with a fieldworker as well as self-recordings. All the data is informal in nature, which is desirable as MLE is not expected to be that present in formal settings. The corpus consists of about 2.4 million words, and the transcripts are compiled from two projects: *Linguistic Innovators* and *Multicultural London English* (Cheshire et al. 2011). The corpus is available through an online corpus analysis software, SketchEngine.

The written language corpus was compiled through Twitter (referred to as the Grime corpus). The corpus consists of tweets from various London-based grime MCs. A maximum of 3,200 tweets was collected from each artist (further information about the data collection can be found in section 3.2). The corpus consists of original tweets from the MCs as well as tweets that they have retweeted (reposting a message of another user). In total, the corpus consists of 60,601 tweets and 597,529 words. This amount is inferior to that of the spoken corpus, but this should not be a problem, as this number of tweets should provide a sufficient amount of data to draw conclusions about the prevalence of MLE features in this context.

Section 2.5 notes that grime (as well as MLE) is very much associated with black males, and to a certain extent, that is true, as many grime MCs do represent that description. However, in order to showcase the diversity of grime artists/MLE speakers the list also includes female artists and artists from various ethnic backgrounds. In total, there are 20 artists, of which 5 are females. The number between genders is not equal, but this is not a problem in terms of the data analysis. Five females should be enough to represent a group in this context, and furthermore, this way the gender distribution of grime is represented more realistically. All Twitter profiles used in the corpus are verified by Twitter. Table 3, at the end of this section, showcases the artists whose tweets are included in the corpus. The table includes names, artist names, Twitter handles, followers, gender,

and the number of tweets used in the study.

Retweets are included in the corpus, which could seem problematic. However, the focus of this study is not on the artists, per se. The choice to use Grime MCs as a data source was primarily because they are known users of MLE features. Therefore, the inclusion of retweets is not an issue, as we can assume that the retweets are messages that the artists want to share, and they could also include features of MLE. The motives for retweeting vary from user to user (Boyd, Golder & Lotan 2010, p. 6). However, according to Marwick and Boyd, "Generally, highly followed users RT or link to items that interest them and presumably their followers" (2011, p. 147). Therefore, we can assume that the reposted tweets provide relevant material for analysis. By including retweets, it is actually possible to get a broader range of users in this study. Retweets were compiled into their own subcorpus, so they were excluded from the study when needed. Furthermore, the inclusion of retweets makes it possible to detect whether features appeared more in the tweets of the artists themselves or in the content they had retweeted. The differences between the retweet subcorpus and the rest of the corpus is tested in section 4.1.2.

Another issue with the two corpora is that the participants are quite different between them. The MLE corpus consists mainly of speech samples from adolescents between the ages 4-18. On the other hand, Grime artists, used in the written corpus, are much older – most of them are in their thirties. The context of the corpora also differs quite considerably. By looking at the tweets of Grime artists, it is expected to see an abundance of music-related content, whereas the content of the MLE corpus is much more general. However, this should not be a problem. This study examines the language used, not the actual content of tweets. One could also argue that MLE and the language that Grime artists use are not the same thing. Nevertheless, as noted in section 2.5, the language of grime is most often seen as MLE, and in this study Grime language is seen as being part of MLE. Even though, general MLE and the language of Grime undoubtedly differ in some ways, previous researchers (e.g. Drummond 2018; Adams 2019) have noted the connection between MLE and Grime. Thus, this study considers "grime lingo" as part of MLE.

Table 3. Artists in the Grime corpus

Name	Known as	Twitter username	Gender	Followers (as of 8.2.2021)	Tweets & replies	Retweets	Total
Rachel Prager	Baby Blue	@IamBabyBlue	Female	19,000	2663	473	3136
Jahmaal Noel Fyffe	Chip	@OfficialChip	Male	616,000	1998	1141	3139
Cleopatra Humphrey	Cleo. / Mz. Bratt	@ItsCleopatra	Female	44,000	2264	821	3085
Darren Dixon	D Double E	@DDoubleE7	Male	110,000	1520	1628	3148
James Devlin	Devlin	@DevlinOfficial	Male	245,000	2426	722	3148
Dylan Kwabena Mills	Dizzee Rascal	@DizzeeRascal	Male	387,000	1854	1262	3116
Deshane Cornwall	Frisco	@BigFris	Male	107,000	2164	999	3163
Justin Clarke	Ghetts	@THEREALGHETTS	Male	156,000	2348	814	3162
Nathaniel Thompson	Giggs	@officialgiggs	Male	360,000	2708	459	3167
Jahmek Power	Jammer	@jammerbbk	Male	117,000	2919	249	3168
Jamie Adenuga	JME	@JmeBBK	Male	981,000	3200	0	3200
Kane Brett Robinson	Kano	@TheRealKano	Male	324,000	454	2538	2992
Louise Harman	Lady Sovereign	@ladysov	Female	59,000	3090	101	3191
Maxwell Owusu Ansah	Lethal Bizzle	@LethalBizzle	Male	248,000	2926	252	3178
Niomi Arleen McLean-Daley	Ms. Dynamite	@Ms_Dynamite	Female	103,000	1199	1478	2677
Kojo Kankam	Novelist	@Novelist	Male	90,000	1044	701	1745
Paris Moore-Williams	P Money	@KingPMoney	Male	132,000	2989	200	3189
Anthony Harris	President T	@Prez_T	Male	15,000	1654	1451	3105
Chanel Cali	Shystie	@IAMSHYSTIE	Female	23,000	2477	215	2692
Joseph Junior Adenuga	Skepta	@Skepta	Male	1,100,000	2164	1036	3200

3.2 Data collection

As noted in section 2.7, the language of social media tends to be informal, and non-standard varieties are very evident in social media. Therefore, social media, more specifically Twitter, was selected as the data source for this study, as it provides recent and accessible data. Once Twitter was selected as the data source, the decision about participants needed to be made. It was challenging to find "regular" MLE users, so attention was turned to public figures who are known users of MLE. The choice to use public figures also decreased some ethical issues, which will be discussed in section 3.3. In section 2.5, Grime artists are described as MLE users, and they even actively contribute to new features of MLE as well. Section 2.5 also notes that grime has spread outside of

London both domestically and internationally, and that Grime MCs are avid users of MLE (or MUBE) elements regardless of their location. However, the purpose of this study is to examine MLE features, so it is logical to limit the scope to just London-based artists. London is the central location for grime and keeping the scope on just London artists enhances the chances to find MLE features.

The definition of a grime artist is a vague one, as many of the artists included have worked with other genres as well. The list of artists was compiled with help from the subreddit r/grime (nikoma, 2015; plain_lou, 2019) and it includes artists identified as Grime musicians by the general public. The aim was to pick out prominent grime artists from London to represent the users of MLE online. It should be noted that some essential names had to be omitted from the list. For example, one of the most known Grime artists, Stormzy, deleted his account at the start of 2020. At the time of deleting, Stormzy had over 1,400,000 followers on Twitter according to SocialBlade. Another big name, Richard Kylea Cowie Jr., better known as Wiley, had to be left out, as his Twitter account was closed in the October of 2020 after a series of controversial tweets.

The tweets were collected on the 22nd of February 2021 using Vicinitas, which is an analysis tool for Twitter (Vicinitas 2021). The program lets one collect 3,200 of the most recent tweets per user, which was sufficient for the purpose of this study (for most users, the software gave little over 3,100 of the most recent tweets). This limitation is also good because had the study included all the tweets from all the users, there would have been over-representation from artist who tweet more than others. Vicinitas transforms the tweets from each user to a single Excel file, from which they were first compiled into text documents, and then compiled into a corpus using SketchEngine. On SketchEngine, the data was divided into different subcorpora. Tweets and replies were compiled into a subcorpus and retweets to their own. The data was also divided into subcorpora according to gender, so it would be possible to see whether gender plays a part in the usage of MLE elements.

The spoken data corpus, Multicultural London English Corpus, is tagged with Penn Treebank tagset v2.5. The tool produces part-of-speech tags (POS-tags) for the corpus. In order to make the two datasets comparable with one another, the Grime corpus was also annotated using the same annotation tool SketchEngine. Information about the tags used and the tool itself can be found in the article by Marcus, Marcinkiewicz, and Santorini (1993) and SketchEngine (2015a). It must be noted that the accuracy of POS-tagging is never 100%, and the accuracy surely drops even more when looking at non-standard varieties of English. However, the POS-tags in this study are mostly used to obtain examples for further analysis, not to draw conclusions from the tags themselves, thus making the inaccuracy not so relevant here.

3.3 Ethics

The line between public and private conversations is often blurred in social media (Ahmed, Bath and, Demartini 2017, p. 6). Therefore, researching the language of social media can be problematic from an ethical standpoint. Twitter is considered to be a public space, which is supported by Twitter's privacy policy document, which states, "Twitter is public and Tweets are immediately viewable and searchable by anyone around the world. We give you non-public ways to communicate on Twitter too, through protected Tweets and Direct Messages. You can also use Twitter under a pseudonym if you prefer not to use your name." (Twitter 2021). However, as Ahmed, Bath and, Demartini (2017, p. 6) note, it is important to consider how aware Twitter users are about the publicity of their tweets. Twitter users have accepted the terms of use, but it still might be unclear to them that their tweets can be viewed by the whole world, and furthermore used by third parties. Terms and conditions are often left unread, and the terms are often very vague and complex to understand, so through them, it is not possible to obtain valid informed consent (Luger, Moran & Rodden 2013, p. 2687).

Most of the data in this study is from public figures, who are probably aware that their tweets are accessible to a large audience. However, a considerable amount of data are retweets which include tweets from "regular users" who might not be as aware of the publicity of their tweets. Usually, in academic research the participants are asked for their consent. However, in this case that is not possible. The inclusion of retweets makes the number of users in the corpus rather vast. Even if the study concentrated on only the artists' tweets, it would be improbable that all of them could be reached to give their consent. Therefore, this study will trust, to a certain extent, that users are aware that their data can be used for academic research. However, examples are only presented from the actual tweets of the artists, not from the retweet subcorpus, to avoid ethical problems.

By collecting and publishing tweets, this project also processes personal data, as users can easily be identified via their tweets. The Finnish National Board on Research Integrity states that, "It is not generally appropriate to publish the data of people who have participated in the research in a way that allows them to be identified. This does not apply to public figures who exercise or who have exercised significant power and whose privacy is narrower than other individuals" (Kohonen, Kuula-Luumi & Spoof 2019, p. 15). Thus, it should be ethically justified to use the tweets from the grime artists, as they do fit this description of public figures. The Grime Corpus is not going to be published, which also protects the privacy of "regular users".

3.4 Methods

As mentioned, all the features presented in section 2.4 are not going to be analyzed. The keyword tool is going to be used to get a general view of the language of the corpus. On top of that, this study also investigates phonologic elements, heavily concentrating on the realization of dental fricatives in written form. MLE's slang terms are also examined as well as the new pronoun innovation *man*. These features were selected for this study because they are noted as important elements in previous research, and it was feasible to search them in the corpus/corpora. Section 3.4.1 explains the use of the keyword tool, and 3.4.2 presents the search parameters for the features that were examined.

3.4.1 Keyword analysis

The analysis of the corpora was done with SketchEngine. For general observations about the language of the Grime corpus this study utilizes the keywords tool. "The keyword tool compares corpora and identifies what is unique or typical. The selected corpus is compared to a reference corpus to identify key data" (SketchEngine 2021a). Through the keyword analysis it was possible to extract the most notable words from the corpora.

The reference corpus that was used when looking at keywords was The English Web Corpus (EnTenTen). EnTenTen is an English corpus collected from various texts from the internet. It offers a comprehensive overview of the language used in online communications and includes almost 22 billion words. "The corpora are built using technology specialized in collecting only linguistically valuable web content" (SketchEngine 2021b). Therefore, it is a good comparison point to seek elements that differ from the general English that appears online. In each keyword analysis, the top 200 words were included in the analysis.

The keyness score for the keywords was calculated via a method called *simple maths*. The formula for the keyness score is presented below (SketchEngine 2015b).

$$\frac{fpm_{rmfocus} + N}{fpm_{rmref} + N}$$

$fpm_{rmfocus}$ is the normalized (per million) frequency of the word in the focus corpus,
fpm_{rmref} is the normalized (per million) frequency of the word in the reference corpus,
N is the so-called smoothing parameter ($N = 1$ is the default value).

The method for analyzing the subcorpora was to compare the both the retweet subcorpus and the tweets and replies subcorpus to the EnTenTen corpus. This allowed to see MLE features from both subcorpora along with their mean frequencies. The mean frequencies of the subcorpora

were then compared to one another to see whether elements were more common in the original tweets or the retweet subcorpus.

3.4.2 Analysis of features

When looking at individual features, this study will analyze frequencies between the two corpora and the collocations (words that co-occur together statistically more often than by chance) of the features. By doing this, it is possible to detect both quantitative and qualitative aspects of the features. The two datasets are different in size, so the frequencies of words were normalized, meaning that the number of occurrences is divided by the total amount of words in the corpus and then multiplied by 1,000,000 to yield the normalized frequency per million words (SketchEngine provides the normalized frequencies by default so there was no need to calculate them). Median frequencies were also used to analyze the features. This was done, to see if features were used by the artists evenly or if the use was limited to only some users. Median frequency is the middle value in an ordered dataset. This means that normalized frequencies for each person were calculated, then the values were ordered by frequency, and then the frequency in the middle (in this study the average of the 10^{th} and 11^{th} value, as there were 20 participants) was the median frequency. This was also automatically available on SketchEngine.

When looking at collocations of the features, this study aims to find fixed expressions and the context in which the tokens appear. The range of collocations was 1 word on either side of the focus word. The statistical significance for the collocations was calculated via their T-score. The T-score expresses the association between the words and ensures that the co-occurrence of the words is not unsystematic.

The list of MLE features was presented in section 2.4. This study will not examine all the features. Some features were omitted because it proved too difficult to develop relevant search criteria for the features. Others were excluded because there were too little or no occurrences of the features. This section introduces the features that are examined in this study, including the search terms.

Dental fricatives

In section 2.4.1, Dh-stopping and Dh-fronting were mentioned along with Th-stopping and Th-fronting as MLE features regarding dental fricatives. The words included in the study are showcased in the results section 4.2, along with discussion about the findings. As a reminder:

Dh-stopping: /ð/ → /d/

Th-stopping: /Θ/ → /t/

Dh-fronting: /ð/→ /v/

Th-fronting: /Θ/ → /v/

Examples of the features were searched via the wordlist tool in SketchEngine. The wordlist tool generates frequency lists of various kinds (words containing certain letters, parts of speech, word forms etc.). Here, the wordlist tool was utilized to search for certain letters in words to find words in which these features are visible. The list of words containing *th* was used with all the features. Words containing *th* were collected, and then the corpus was searched for words that would replace the *th* with the letters *d, t, v,* and *f* (e.g. The word *thing* appeared in the wordlist of words containing *th*. Then *ding, ting, ving* and *fing* were searched in the corpus). As well as that, the wordlists for letters *d, t, v,* and *f* were looked at to find more examples of words where *th* was replaced by the aforementioned letters.

Use of MLE slang words

The list of MLE slang words for this study was compiled with the help of Chris Nott's MLE glossary (2013). The words of the glossary were supplemented with slang terms from previous researchers. The list of slang words that were included in this study can be found in section 4.3.

Mean frequencies of the slang words were calculated for both the MLE corpus and the Grime corpus. Then the frequencies were compared to one another to see whether words would appear more often in written or spoken mediums. As the list is rather long, the analysis was primarily quantitative, but some key differences regarding slang vocabulary between the two corpora was analyzed through examples. It should be noted that some MLE slang terms had to be omitted from the list because they are homonyms of other common words. Examples of omitted slang terms are *long* meaning "an expression of annoyance or something that requires a lot of effort", *yard* referring to "one's area or home" and *peak* as a "replacement for many negative adjectives" (Urban Dictionary[1]).

Man pronoun

This study looks at the properties of *man* to see if it is used to replace some pronouns more than others. Looking at *man* as a pronoun was slightly problematic. In the MLE corpus there are 2,672 instances of the word and in the Grime corpus the word appears 2,249 times. It is not possible to classify when *man* is used as a pronoun with machine-driven tools as annotation tools do not

[1] Urban Dictionary is used for the definition of some of the slang terms in this study with due caution. For some of the slang terms, there are no official definitions available outside of Urban Dictionary, and thus it was used in this study. Urban Dictionary is not as reliable as traditional dictionaries, but it does provide accessible and up-to-date data for new language innovations.

recognize *man* as a pronoun. It is not feasible to search through all instances within the scope of this MA thesis. Therefore, a sample of 200 instances from the Grime corpus was randomly selected, which was manually analyzed. Examples of *man* were retrieved from SketchEngine, and then transferred to a text file. Then the examples were analyzed to find out which person pronoun was *man* taking the place of. This was not necessary for the MLE corpus, as the results regarding *man* as a pronoun are already recorded in the study of Hall (2020, p. 122). Those findings are presented in section 4.4.

Section 4.4 investigates how often *man* was used to replace a personal pronoun, and which personal pronouns did *man* most often replace. Another thing that was examined regarding *man* is whether gender affects the usage of *man* as a pronoun. Section 4.4.2 inspects how often women use the new pronoun and whether we can detect any differing patterns compared to the males in this study.

4. Results

The results are divided into subsections. The first section 4.1 showcases the results from keyword analysis. Section 4.1 also includes a comparison between the retweet subcorpus and the original tweets subcorpus. Section 4.2 views the phonological features in written form, mainly concentrating on the realization of dental fricatives. 4.3 explores the findings regarding MLE's slang lexicon, and 4.4 examines the results regarding MLE's new pronoun innovation *man*. Each subsection here also includes analysis of the features, which will be elaborated in the discussion section 5.

4.1 Keywords

The first section of the keyword analysis explores the keywords when comparing the whole Grime corpus (including the retweets) to the English Web 2018 (EnTenTen) Corpus. The second section 4.1.2 compares the frequencies of keywords (EnTenTen still being the reference corpus here) from the retweet subcorpus to the frequencies in the subcorpus including only the original tweets from artists.

As explained in section 3.4, keyword analysis showcases what is typical for a corpus compared to a reference corpus. The aim of using keyword analysis was to see general observations about the corpora. This section showcases the findings about the language of the corpora, concentrating on the MLE elements. The analysis of features will be extended in the sections about the individual features. The top 200 keywords were used in analysis, which can be found in the appendix.

4.1.1 Grime corpus compared to the English Web 2018 corpus

The first keyword analysis was the comparison of the whole Grime corpus to The EnTenTen corpus. The whole corpus was used, including retweets, to get a complete picture of the language used in the corpus.

It was no surprise that many of the keywords from the Grime corpus were music-related words. Artist names, songs, and albums appeared very frequently on the keyword list. This was to be expected, as artists frequently use Twitter to promote their music. 7 of the top 10 keywords were music-related words (2. Skepta, 3. Grime, 4. Ghetts, 5. Frisco, 7. Giggs, 8. Kano, 10. Dizzee) and in general music-related words were very present in the keyword analysis.

Another visible theme of the keywords was the use of internet slang, which like music-related words, was no surprise. For example, *lol* ("laughing out loud") appeared many times, with differing amounts of o-vowels (6. *lool,* 17. *looool,* 22. *loooool,* 23. *lol*). Other common internet slang abbreviations that were found are visible in the table below (Table 4). The use of informal language is very present on Twitter, whereas the EnTenTen corpus includes a broad overview of the language of the internet, thus making these abbreviations and other internet slang-related words frequent in the keyword analysis.

Table 4. Internet abbreviations from the keyword list

#25 icymi (in case you missed it)
#43 ffs (for fuck's sake)
#50 tbh (to be honest)
#59 ppl (people)
#62 tbf (to be fair)
#101 wtf (what the fuck)
#140 lmao (laughing my ass of)

The third category of words, and the one that is particularly interesting in terms of this study, was MLE elements. For example, *ting* was the 13th item on the list with 517 instances on the corpus, 574.0 instances per million tokens (ipm). This would suggest that Th/Dh-stopping, a prevalent feature of MLE, has spread to written form as well with MLE speakers. This notion is supported by the fact that there were several other words in the keyword list, where dental fricatives were realized with the letter *d*:

55. *Dem* (non-standard spelling of *them*), 253 occurrences (280.9 ipm)

85. *Wid* (non-standard spelling of *with*), 63 occurrences (70.0 ipm)

106. *Dat* (non-standard spelling of *that*), 134 occurrences (148.8 ipm)

161. *Dere* (non-standard spelling of *there*), 35 occurrences (38.9 ipm)

Therefore, we can assume that phonological features have spread to written context as well, at least in the case of dental fricatives. The phenomenon of Th-/Dh-stopping and Th/Dh-fronting is discussed in more depth in section 4.2.

Th/Dh-stopping was not the only MLE feature that was found in the keyword analysis. Address and reference terms that are linked to MLE were also prevalent in the analysis of keywords. Adams' study (2018) explored the address terms used in grime artists' speech and links the use of these address terms to MLE.

Fam (short for *family* but can refer to close friends as well [Urban Dictionary]) was the 14th item in the keyword analysis, the highest position for a reference term. Adams' study (2018, p. 17) notes *fam* as the most common address term with Grime MCs, which is supported from the data here as well. Examples 1 and 2 showcase the use of the word in the Grime corpus. *Fam* appeared 404 times (448.6 ipm) in the grime corpus but only 15 times (5.1 ipm) in the MLE corpus. This would suggest that the address term is more linked to the language of Grime than it is to MLE. Furthermore, the use of *fam* in the MLE corpus was exclusively limited to using the words as a abbreviation of *family* whereas in the Grime corpus the term was mostly used as an address/reference term (Examples 1 and 2).

(1) @Novelist Mans good good bars **fam** 🤣🤣(@BigFris, 2020-02-09, 1226496515717390337)[2]

(2) @user yea mate its good to be back good morning all, hope everyone had a good friday night mans on a cheese omlette hype right about now, say summin **fam** , wat (@DevlinOfficial, 2015-03-20, 1144365319152132097)

Bruv was 18th on the list with 150 occurrences (166.5 ipm). According to the keyword analysis, *bruv* seems to be a very common address term in the context of grime. It was also very common in the MLE corpus with 390 occurrences (131.8 ipm), so it seems that it is very much an MLE feature. The word *bruv* also gives evidence of Dh-fronting, as the word is a shortened version of *bruvver*/*bruvva*. It is one of the many address terms derived from the word *brother*. Other derivations that were visible in the keyword analysis were *brudda* at 30th with 107 instances (118.2 ipm), *bruva* with 37 hits (41.1 ipm), *bredrin* with 25 occurrences (27.8 ipm), and the more general

[2] Examples of this study are all presented in the same way. Tweets are displayed in full if not otherwise stated (this is done with square brackets […]). Bolded words are added to highlight relevant words. Mentions of users outside the 20 artists of this study are anonymized (@user), for ethical reasons. The information after the tweet, in brackets, consist of the username, date (YYYY-MM-DD), and ID of the tweet.

bro with 990 occurrences (1,099.17 ipm) at 9th place. *Bro* is a very common slang term, and it is used widely in many non-standard Englishes; therefore it is not an MLE-specific term. However, it seems that the word is also very common in the context of Grime. In the MLE corpus, the address term *bredrin* (although spelled bredren) was the second most common derivation of brother with 42 instances (14.2 ipm). Other derivations were non-existent or very uncommon in the MLE corpus.

Mate was another popular address term, at 157th on the list, with 533 hits (591.8 ipm). It was also prevalent in the MLE corpus, with 1,114 instances (386.6 ipm). *Mate* is commonly used in informal British English, and it seems that MLE speakers have adopted the word to the feature pool. *Hun* (short for honey [Cambridge Dictionary]) was also used as an address term commonly, appearing at 81st place with 105 instances (116.6 ipm). However, it should be noted that the term was almost exclusively used by women, with only 2 instances from a male user. These terms are obviously not MLE specific, but it showcases the fact that British slangs (such as Cockney) have affected MLE.

The keyword analysis also included many lexical items of MLE. Number 58. was *mandem* with 60 results. According to the Macmillan Dictionary *mandem* is "a group of friends, especially male friends". It was relatively rare in the MLE corpus, with only eight instances found in the corpus. However, it is noted as an MLE feature in several publications (e.g. Kerswill 2014 & Drummond 2018).

Nang, meaning "Cool, good, a positive" (Nott 2013, p. 2) was another slang term that appeared in the analysis at 104th place with 58 instances. The slang term *gyal*, which is the way speakers of Caribbean Creoles say *girl* (Urban Dictionary), was 138th on the list with 34 instances. Other slang terms that appeared in the analysis of keywords were *skint* (poor [Urban Dictionary]), at 107th, *yute* (person of a younger generation, a young child/baby [Nott 2013, p. 83]) at 179th and endz (your area, home or zone [Urban Dictionary]) at 180th. As noted in section 2.4.2, many slang terms of MLE are from Jamaican Patois, and this point seems to be evidently clear at least in the context of Grime. Section 4.3 will delve deeper into the analysis of slang lexicon.

The distinction of linguistic elements specific to grime and MLE lexicon can be ambiguous. Therefore analysis of some of the features is problematic since it is not clear whether the elements should be considered as being part of MLE or "grime language". This issue was apparent, especially when looking at terms related to music. For example, the 12th item on the list *banger* (A great song [Urban Dictionary]) and number 19 *riddim* (rhythm [Urban Dictionary]) were commonplace in the Grime corpus. The aforementioned terms did not appear in the MLE corpus, and thus they could be considered more grime-specific terms.

Number 34 on the list *coz* was an interesting finding because it is not noted as being a

feature of MLE in previous research. However, according to the keyword analysis, it seems to be used very commonly with MLE speakers. It is a shortened way to say *because* (Urban Dictionary). With the spelling *coz* the word appeared 177 times and as *cos* it was found in 79 instances (in total 256 occurrences). The standard *because* appeared 258 times in the corpus, so it is clear that grime artists often replace the standard spelling with either *coz* or *cos*. This phenomenon is also noted in the MLE corpus, as *cos* appears 9,993 times, whereas the standard *because* is much more uncommon with 4,671 occurrences. Therefore, it seems to be a very common feature of MLE, not previously noted in research.

The discourse marker *innit* was noted in section 2.4.2 as a common feature of modern British English, especially amongst young speakers from the London area. Surprisingly, the term appeared at 139th with only 35 instances. Furthermore, 24 of the 35 instances were from two MCs, Chip with 13 and JME with 11. Therefore, it seems that *innit* is not used very frequently, at least in the context of grime. Or perhaps its use on Twitter, in general, is more uncommon than it is in real-life conversations. However, *innit* was higher up on the list at 112th place with 80 occurrences. In the MLE corpus, *innit* appears a whopping 3,181 times (1,074.9 ipm). The total mean frequency of *innit* and *init* was 127.7 ipm, so it is clear that the use of the term was much more widespread in the MLE corpus. Therefore, it seems that the term has not spread to written language to the extent to which it appears in spoken mediums. We must also consider the fact the MLE corpus is rather outdated as it is ten years old at the time of writing, so the feature might be on the decline. However, Martinez's study (2014, p. 384) notes that *innit* is a pervasive feature of London teenagers. It must also be considered that perhaps *innit* is not used so much in the context of Grime as it is with more general speakers of MLE. However, the important notion here is the fact that *innit* has spread to written context, to some extent, but it is not anywhere near as common as it is in the MLE corpus. Furthermore, at least grime artists opt for the use of the spelling *init* rather than *innit*, which was a surprising innovation. Examples 3 and 4 present the use of *init* in the Grime Corpus.

(3) 😂 I'm an old codger **init**. But I'm still alive and still relevant 20 years later this is a great birthday gift 🎁 https://t.co/AYQF8stKZE (@jammerbbk, 2019-07-02, 1146184428500246528)

(4) Took my son to see Sonic today, no sorry let me rephrase that.... I brought my son with. he'd already seen it but I needed to see it **init** 🤣🤣 (@KingPMoney, 2020-02-20, 1230618576303857664)

4.1.2 Comparison of subcorpora

The inclusion of retweets was discussed in 3.1. It was hypothesized that the inclusion of retweets would not be problematic and that retweets would be beneficial in order to get a more extensive scope for the study. This hypothesis was tested here by comparing the results from the retweets subcorpus to the subcorpus that only includes original tweets.

This section demonstrates the MLE features found in the keywords of the retweet subcorpus and contrasts the mean frequencies of them to those in the original tweets and replies by the Grime artists. Here again, EnTenTen was used as the reference corpus. Table 5 showcases the normalized frequencies of MLE elements in the subcorpora. The items on Table 5 were collected from the top 200 keywords of the keyword analysis of the retweet corpus.

Table 5. Normalized frequencies of MLE elements in the subcorpora of the Grime corpus

	Retweet subcorpus	**Tweets & replies**
23. Ting	574.0	715.3
65. Brudda	54.7	159
70. Dem	207.5	326.9
73. Fam	109.5	661.1
84. Bruv	43.2	243.8
97. Endz	37.5	23.5
109. Mandem	34.6	86.7
114. Da	495.6	424.5
144. Gyal	28.8	43.4
149. Wid	37.5	90.3
192. Dat	74.9	195.1

When comparing the results from the two subcorpora, it is clear that messages that the artists have retweeted also contain many MLE elements. However, the mean frequency is higher in the original tweets subcorpus with all terms, barring two items *da* and *endz*. It should also be noted that many MLE elements that appeared in the keyword analysis of the whole corpus, were not present here. Therefore, it seems clear that the use of MLE is more evident in the original messages of the artists rather than in the content that they had retweeted. However, there were plenty of MLE elements in the retweet subcorpus, thus confirming that the retweet corpus does provide significant data in this context, although MLE elements were generally rarer there.

4.2 Phonological elements

The analysis of keywords revealed that the Grime artists express dental fricatives in a non-standard way in written form. This section will delve deeper into the phenomenon of Th/Dh-stopping as well as Th/Dh-fronting. The focus is heavily on the representation of dental fricatives, but some notions about other phonological features in written form will be presented at the end of this section.

The phonological elements studied here are part of many other non-standard varieties of English as well. For example, Th-stopping is used by AAVE speakers (Eisenstein 2013, p. 15) and Th-fronting is a feature of Cockney (Schleef & Ramsammy 2013, p. 26). Therefore retweets were excluded in the analysis of phonological elements, as it would be possible, although not likely, that messages that the artists had retweeted would be written in AAVE for example, thus falsifying the results.

When excluding the retweets, the number of tweets per user changes greatly. At this point, an individual user could severely distort the results, so instead of looking at just normalized frequencies, this study also takes median frequencies into account. As explained in 3.4.2, this allows seeing if the features are used by the artists equally or if the use is limited to only some users.

Comparing the data from the Grime corpus to the MLE corpus regarding phonological elements was problematic. The transcription of the MLE corpus is mostly orthographic, not phonemic, so it does not consider all the pronunciational elements of the speakers. For example, there were little to no instances of Dh/Th-stopping visible in the MLE corpus, although both phenomena are recorded features of MLE (Drummond 2018, p. 173). Therefore the data from the MLE corpus is not comparable in terms of phonological elements (with most items here at least).

4.2.1 Dh-stopping

The results of Dh-stopping are presented below. Dh-stopping can occur anywhere in the word (Kerswill 2018, p. 172). Using the wordlist tool, it was possible to seek out words that start with the letter *d* and then select the instances where *d* was used to replace the standard *Th* spelling to find out words where Dh-stopping happens in an initial position. Similarly, the wordlist tool was used to detect words where Dh-stopping would appear in medial and final positions. The items that were picked out from the wordlist tool can be found in Table 6 below, along with mean frequencies and medians.

Table 6. Dh-stopping in the Grime corpus

Non-standard spelling			**Standard spelling**		
Word	Mean frequency	Median frequency	Word	Mean frequency	Median frequency
Da	424.47	120.1	The	19,435.3	20,549.5
Dem	280.9	200.4	Them	1,081.9	801.9
Dat	195.08	71.0	That	6,032.9	6,399.5
Di	61.41	18.2	The	19,435.3	20,549.5
Dere	43.35	0	There	1,638.3	1,439.1
Dis	52.38	29.9	This	6,759.0	6,434.4
Wid	90.31	0	With	3433.7	4,125.4
Brudda	158.95	18.2	Brother	700.8	347.8

Table 6 above presents the findings regarding Dh-stopping. The table includes the most frequent instances of Dh-stopping in the corpus, but it should be noted that they were not the only ones where Dh-stopping was visible. However, other examples were more uncommon to the point that they cannot be considered as widespread innovations but rather expressions of an individual user.

When looking at the normalized frequencies of the tokens, it seems like Dh-stopping is relatively established. However, the normalized medians showcase a more realistic view of the phenomena. In many cases, Dh-stopping was very prevalent with a couple of the users, while with others, it was uncommon or non-existent. In none of the cases did the non-standard spelling occur with every user, although some tokens got pretty close (*dem* 16/20, *dat* 16/20, and *da* 13/20 users).

The word *dem* seems to be the most frequent word where Dh-stopping occurs when comparing the number of the word to that of the standard spelling. However, it must be noted that in some cases, the lemma *dem* was used as part of a name of an album or a song (e.g. Giggs' album "Wamp 2 Dem" and D Double E's song "Dem Tings Dere"). *Dem* can also be used to indicate plurality (e.g. Man dem, gyal dem), which was also visible in the results. However, most of the results were relevant occurrences where the word was used to replace the spelling of *them* (Examples 5 and 6). The median frequency was slightly lower than the mean frequency, but *dem* appeared in 16 out of the 20 artists' tweets, which supports the claim that the non-standard spelling is rather established here.

(5) LET **DEM** KNOW MY BRUDDA. […] (@officialgiggs, 2018-07-25, 1022168233464487936)

(6). Then funniest thing is people from SW say their from South. People from NW don't say their from N. Don't let me start adding NW man into the team... it will get mad long for everyone. NINES, SKRAPZ... r u mad???!!! **Dem** man deh some proper rappers bruv!!!! No wish washy lines. (@OfficialChip, 2019-04-30, 1123232458164584449)

The definite article *the* was also a subject of Dh-stopping. There were many different spellings evident in the corpus, with *da* being the most frequent one with 235 occurrences. The second most common way to replace *the*, that involved Dh-stopping was the spelling *di* that appeared 34 times in the corpus. *De* was also used a handful of times, but it was too infrequent to be considered as an established feature. There was a rather vast gap between the median and mean frequency in terms of *da*. Giggs and Frisco, with mean frequencies of 3,300.2 and 3,118.2, respectively, were by far the most avid users of the word, with others far behind them. However, the median frequency is still high enough in order for the spelling to be regarded as somewhat established.

Dh-stopping was also present in the word *there*. The spelling *dere* appeared 35 times, but in many cases, it was used to refer to D Double E's song "Dem Tings Dere". *Deh*, as a replacement for *there,* appeared 29 times in the corpus, of which 19 times it can be counted as being relevant (In 10 cases it was used to talk about a song called "Deh Deh"). *Deya* was also used eight times (14.5 ipm). So, in terms of *there*, it cannot be said with certainty that there would be anything systematic going on, although some users here opt to utilize Dh-stopping with the word.

The word *den* appeared to be quite prevalent when looking at just the frequency of the word in the corpus (138.8 ipm). However, *den* was only used a handful of times to replace the word *then,* and in most cases, it was used to refer to a Grime event called "The Den". *Dis* also appeared quite frequently in the corpus. 15 of the 29 instances were by Giggs, although *dis* was prevalent in the data of 10 other artists as well. However, *dis* was also used to refer to somebody and as a shortened version of *disrespect.* Thus, the results of *dis* are not regarded very notable here.

Brudda as a replacement for *brother* was used 88 times (159.0 ipm). That was the only instance where Dh-stopping was used somewhat systematically in a medial position. Other instances, such as *mudda* for *mother* used only 5 times, were too rare in order to be regarded as significant findings. *Bredda* was another spelling for *brother* where Dh-stopping was visible, but it only appeared 11 times (19.9 ipm), although it adds to phenomenon of Dh-stopped brother.

It was clear that Dh-stopping occurs mainly in word-initial positions. Instances of Dh-stopping were also searched in medial and word-final positions, but it was clear that Dh-stopping was almost exclusively not used in those contexts. By looking at the first 500 results of the wordlist

that contained the most frequent words ending with the letter *d*, only one item was found where Dh-stopping was used. That word was *wid*, that appeared 63 times in the corpus.

As explained, in some cases, Dh-stopping was used when talking about a particular musical piece, event etc. This does skew the results slightly, but there are still enough results from "natural incidences" to prove that in the context of grime Dh-stopping does frequently appear in written form as well, at least with three words, *Da*, *Dem*, and *Dat*. These words seem to be well established in the language of Grime artists, and Dh-stopping, in general, appeared in many words in the corpus, thus suggesting that the phenomenon would have spread to written form to some extent at least.

4.2.2 Th-stopping

Similarly to Dh-stopping, the list for Th-stopping was compiled with the help of the wordlist tool in SketchEngine. The words including *th* in them were collected, and then terms that would replace *th* with just the letter *t* were searched in the corpus. The ones that had more than 5 results in the corpus are included in the table below. However, as visible in Table 7, there were not many instances of Th-stopping in the corpus.

Table 7. Th-stopping in the Grime corpus

Non-standard spelling			**Standard spelling**		
Word	**Mean frequency**	**Median frequency**	**Word**	**Mean frequency**	**Median frequency**
Wit	7.2	0	With	3433.7	4,125.4
Ting	715.3	637.2	Thing	830.9	643.4
Yout	32.5	0	Youth	34.3	0

The one item that stands out in terms of Th-stopping is *ting*. The spelling was used almost as frequently as the standard spelling *thing*. *Ting* was used by 19 out of the 20 artists, further proving that it is a very common expression in the context of Grime. Drummond suggests that *ting* could be a lexical variant (rather than a phonological one) for the word *thing* (2018, p. 192), and according to the data here, it seems to be quite lexicalized. It is widely used to replace the standard spelling and it can also also "emphasise something or act as a replacement for an object or action" (Nott 2013, p. 13). Examples 7 and 8 show the use of *ting* in the Grime corpus. On top of that *ting* is also used as part of several expressions, which were also present in the corpus. When looking at the collocations of the word, the expression *mad ting* (Example 9) appeared most frequently with 25 instances. The top definition on Urban Dictionary explains *mad ting* as "An expression of surprise, amazement, pleasure, unexpectedness; or, agreement with sentiments thereof". Another common

phrases were *big man ting* (honestly or in all seriousness [Urban Dictionary]) *joke ting* (bad situation [Urban Dictionary]).

(7) I am 20 seasons in this music game. That's 20 years. Please don't compare my **ting** to the new generation. It's there time now. I hope they're still here in 20 years time too. #iStillGotMoreWorkToDo (@LethalBizzle, 2021-02-13, 1360717507473080326)

(8) Love bruva! Check the new **ting** #B2DL5 (@BigFris, 2018-04-02, 980941018961989633)

(9) Looking forward to the New Year it's gonna be a **MAD TING**! 🎆👊🏾💥 (@Prez_T, 2016-12-31, 815138392916815872)

Yout was another element, which can be counted as relevant. The median frequency for *yout* was 0, which would suggest that the word would not be established. However, when looking at the mean frequencies in the corpus, it is possible to detect that the non-standard spelling is as common as the standard one. The word is relatively rare in the corpus, thus making the median not reliable here. Both *yout* and *youth* were used by 9 out of the 20 artists, thus making the non-standard spelling as common as *youth* in that sense as well. JME, Skepta, Novelist and Dizzee Rascal used both terms, and here the context seemed to matter considerably. The standard spelling was used in more official contexts (Example 10), whereas the non-standard spelling appeared in more casual messages (Example 11). It should also be noted that the spelling *yute* for *youth* was also almost as common in the corpus, with a mean frequency of 30.0. If both Th-stopped variants *yute* and *yout* are counted together (62.5 ipm), it is clear that the non-standard spelling for *youth* is actually more common than the standard *youth*.

(10) @user I spread awareness on all parties, I wanted the **youth** to register to vote for any party, the picture you've used is from when i interviewed Corbyn for iD magazine, I'd interview any MP. bbk edition 1 for you. And edition 3. Seckle your jaw. (@JmeBBK, 2020-05-03, 1256900186510630912)

(11) @JmeBBK I thought it was a **yout** app, for upper body dance moves, just arms and neck. Musically ting. (@JmeBBK 2020-12-05, 1335216461913186304)

Besides *ting* and *yout*, there were not all that many instances of Th-stopping in the corpus, which would suggest that in written form, Th-stopping is not that established. There were copious amounts of words where Th-stopping could have been used, but clearly the phenomenon is not very frequent in the data of this study. The two terms *ting* and *yout* seem to be the ones that have

established a non-standard spelling here. Examples of other uses of the phenomenon are presented in examples 12 and 13.

(12) When last did u wake up n tell your family you love them? Bet you woke up and checked ya IG before u brushed ya **teet** doe... mentul. (@OfficialChip, 2018-10-10, 1049951224739651584)

(13) @user @user Mateee. I ain't **wit** them crowded pool parties fuck that I'm sunbathing (@IAMSHYSTIE, 2019-05-29, 1133790549998219267)

4.2.3 Th-fronting and Dh-fronting

Unlike, Dh-/Th-stopping, the discussion about Th- and Dh-fronting is not divided into two sections, mainly due to the fact the phenomena were rather rare. Wordlists containing the letters *f* and *v* were inspected, to find instances where Th-fronting would have been visible. However, instances of the phenomena were not found very frequently at all in the corpus.

Bruv was already mentioned in the keyword analysis as a possible candidate of Dh-fronting. However it was the only instance where Th-fronting was visible to a significant extent. The mean frequency of the word was 243.8 with a median frequency of 51.5. The spelling *bruva* seemed to be quite frequent as well, with a mean frequency of 65.0 but on closer inspection it was found that 34 out of the 36 instances were from Devlin. Furthermore, *bruv* (and *bruva*) could be counted as a lexical variant rather than phonological one.

Interestingly, there were no results in terms of Th-fronting (/Θ/ realized as /f/), although Dh-fronting (/ð/ realized as /v/) was also very rare. According to the findings, or rather the lack of findings, it seems that Dh- and Th-fronting have not been established as features in written form, at least in this context. This might be affected by the participants of this study. Th- and Dh-fronting are traditional features of non-standard British English (Kerswill 2007, p. 6), whereas most participants in this study have a Caribbean or African background, where these features are not used widely.

4.3 Slang words

A list of MLE-specific slang words can be found in this section. Table 5 includes the slang terms used in this paper. The table includes the definition of the words (along with the reference) and mean frequencies from both corpora. Retweets were included in this section in order to get a larger number of results. Individual slang words will be discussed in this section, and the discussion about slang words, in general, will be extended in section 5.2.

Table 8 MLE slang words in the Grime corpus[3]

Word	Source	Definition	Mean frequency Grime corpus	Mean frequency MLE corpus
Batty	Cheshire	Arse or homosexual (UD)	13.3	9.8
Buff	Goldbeck	A person with big muscles (UD)	21.1	14.9
Butters	MLEG	Ugly, horrible, disgusting (MLEG)	5.6	6.4
Crep / Creps	BBC	Shoes, usually sneakers (BBC)	8.9	3.0
Dutty	MLEG	Dirty, horrible, disgusting (MLEG)	13.3	0
Ends/Endz	Kerswill	One's area or neighborhood (BBC)	97.7*	54.7*
Galdem/Gyal dem	BBC	Group of girls (BBC)	12.1	0
Gassed	Goldbeck	Excited or happy (UD)	122.1*	0*
Gyal	-	Girl (UD)	37.8	0
Jheeze	MLEG	Oh my god, surprise, amazement, or substitute for positive acknowledgement (MLEG)	57.7	0
Leng	MLEG	Gun, to be killed or attractive, anything positive (MLEG)	15.5	0
Mandem / Man dem	BBC	Bunch of boys or men, particularly your own group of mates (BBC)	101.0	2.7 + 0.68
Manor	-	Same as ends/endz (UD)	(159.9)	(15.5)
Murk / Merk	MLEG	Attack (physically or verbally), kill, completely intoxicated, to finish off (MLEG)	11.1	0
Nang	MLEG	Cool, good, a positive (MLEG)	64.4	15.9
Par	MLEG	"Breach of social standard / Getting demoralized (UD)	7.8*	0*
Peng	MLEG	Attractive, high quality, anything Positive (MLEG)	15.5	7.8
Rah	MLEG	To express bad, unbelievable, excited, shocking (MLEG)	60.0	3.7
Skank	MLEG	Dance (MLEG)	31.1*	0*
Sket	MLEC	A Prostitute (UD)	6.7	11.2
Skeng		A weapon (UD)	12.2	0
Vex		Annoyed / Irritated (UD)	37.6	5.7
Wavey	MLEG	Drunk (MLEG)	15.5	0
Zoot	Goldbeck	Marihuana (UD)	23.3	6.8

[3]Explanations of the table: *=irrelevant instances omitted, ()= includes irrelevant instances
BBC= BBC 2018. *BBC Radio 4 - Radio 4 in Four - 17 Multicultural London English words and what they mean*.
Cheshire= Cheshire, J., 2013. *Grammaticalisation in social context: The emergence of a new English pronoun*.
Goldbeck= Goldbeck, J., 2018. *What is MLE, who speaks it, and is it safe?*
Kerswill= Kerswill, P., 2013a. *Identity, ethnicity and place: the construction of youth language in London.*
MLEC= Cheshire et al. 2011. *Multicultural London English Corpus.*
MLEG= Nott, C. 2013. *Multiethnic London English Glossary.*
UD= *Urban Dictionary*

It should be noted that the list does not offer a comprehensive overview of all slang terms in MLE. The list is merely a representation of the slang terms that appeared in the corpus/corpora. Several terms that were noted to be part of MLE's lexicon did not appear at all or appeared very infrequently in the corpora, and thus were excluded from the list (e.g. *hench, cotch ,*and many others). It should also be noted that many address/reference terms could have been included in the list. However, they were already covered in the keyword section 4.1, so they were excluded from analysis here even though they could very well be counted as slang lexicon (like they are in many other papers about MLE).

Some slang terms were already briefly discussed in the keyword analysis section. Here again, the distinction between MLE elements and Grime language was slightly problematic. Many terms that appeared in the sources that listed MLE words did not appear in the MLE corpus. The purpose of this study was to detect MLE elements in written form. Thus, it must be considered that some tokens here might be more Grime specific than they are of MLE. However, almost all the slang terms here, barring a couple, were more common in the Grime corpus, so we must also note that the contexts of the corpora also differ. Perhaps the use of slang on Twitter, especially when conversing in the context of Grime, is more common than it is on interviews, such as in the MLE corpus.

As mentioned, the usage of slang terms in the MLE corpus was surprisingly scarce. The fact that address terms were also omitted from the analysis affected this notion. Many of the slang terms found in the keyword analysis of the MLE corpus were address terms, with other slang terms being far more uncommon. There were not many words here that appeared with similar amounts in both corpora. *Buff, butters ,*and *peng* were examples of words that had pretty similar mean frequencies of occurrences in both corpora. *Nang* was another term that appeared to be quite common in both corpora. However, it should be noted that most of the tweets that included the word were about a song called "Nang" by JME and Skepta. In the context described above, the word was used more scarcely, and therefore, it cannot be regarded as a prevalent feature, at least in this context. In the MLE corpus *nang* appeared 47 times. In many instances, it appeared in a metalinguistic context; people discussing the meaning of *nang* and whether it was a common expression or not (Example 14).

(14) yeah to mean sort of it's like it's good like if you go up oh good if you yeah it means good it's like "oh you know what that's **nang**" (MLE corpus)

There were two slang terms that referred to "one's area or neighborhood"; *manor* and *ends* (sometimes spelled with *z* instead of an *s*, *endz*). The results for *manor* are displayed in brackets

because the use word often appeared in instances where it cannot be counted as relevant. In the Grime corpus the word appeared mostly to refer to Kano's album "Made in the Manor" as well as his song "T-shirt Weather in the Manor", though there were 13 instances where *manor* was used "naturally" as well (Example 15). In the MLE corpus *manor* was mostly used to refer to certain locations such as "Manor Park" and "Manor House", although the term also appeared 12 times in its slang meaning too. *Ends* appeared with a mean frequency of 68.8 in the Grime corpus, whereas *endz* had the mean frequency of 28.9, and the terms appeared in 17 out of the 20 artists' tweets, making the term(s) widely used. *Ends* was also prominent in the MLE corpus, although in this case as well, it was used more in the Grime corpus than in the MLE corpus. The mean frequencies here are not exactly correct, because the word also appeared in its non-slang meaning (e.g. as a verb, as in *something ends*), although this was relatively rare. Example (16) shows the use of *ends* in the Grime corpus and (17) in the MLE corpus.

(15) It was nice to be back in the **manor** this weekend to hand out end of season football awards to … https://t.co/TZbYh26fGK (@DizzeeRascal, 2017-05-21, 866227841721094144)

(16) If I go to the **ends** today mandem mite rush me 😂😂😂 coz of young thug!!! I'm vex hahahah (@jammerbbk, 2021-01-05, 1346426544470810626)

(17) you see someone in from a different **ends** shotting a drug. (MLE corpus)

Exclamations *rah* and *jheeze* were common in the Grime corpus, whereas their usage in the MLE corpus was almost non-existent. *Rah* had 11 instances in the MLE corpus, while *jheeze* did not appear at all (transcription of the corpus might affect this as well). In the Grime corpus, *rah* often appeared on its own just as a reaction to a tweet, picture, video etc. When used as part of a sentence, the main use was to emphasize something (Examples 18 and 19). Interestingly *jheeze* was only found at the start of sentences, whereas *rah* appears to be used more flexibly, occurring at the start of sentences medially and in final positions. *Rah* was also used by 10 of the artists whereas *jheeze* was used by 7 (although majority of the instances were from Skepta and P Money).

(18) I said **rah** I won't be going to dance again. (@THEREALGHETTS, 2020-08-28, 1299312738259398657

(19) 17 hours I've just slept....**rah**. (@ladysov, 2014-02-17, 435394683582578688)

There were several other terms as well that frequently appeared in the Grime corpus but had very little to no instances in the MLE corpus, such as *skank* and *gassed*. Both words had several meaning, so the results had to be manually checked to see where it was used in the correct slang

meaning. *Skank* can also mean "a derogatory term for a female" (Urban Dictionary) and *gassed* can stand for the standard language use and the basic form *gas* can also mean "talking rubbish, untrue or false bravado". *Wavey, skeng, leng* and *dutty* were other words that did not appear at all in the MLE corpus but had instances in the Grime corpus.

From the data presented above, it is clear that MLE slang words have spread to written form. It is not all that surprising, since the context of grime supports the use of slang terms and the language of Twitter tends to be colloquial. However, the fact that the use of MLE slang terms was overwhelmingly more common in the Grime corpus was surprising.

4.4 *Man* pronoun

Hall (2020, p.128) notes that the referent of *man* is always context-bound. On Twitter, the context can often be unclear; thus interpreting some of the results for m*an* was problematic. As mentioned in 3.4.2, a sample of 200 instances was randomly selected from the Grime corpus to see how often *man* was used as a pronoun and which pronouns does it most often replace. Section 4.4.2 investigates how women in the Grime corpus utilize this new pronoun. Retweets were excluded in the analysis of *man* because the focus is qualitative here, and enough data can be retrieved from the actual tweets of the artists. Furthermore, as explained in section 3.3, examples from the retweet subcorpus were not used due to ethical reasons, so it was logical to only use the original tweets from the artists here.

4.4.1 Sample

A sample of 200 instances of *man* was randomly selected from the Grime corpus. The sample was analyzed in terms, how many times *man* was used as a pronoun and which pronoun *man* most often replaced. Table 9 below shows the findings regarding the pronoun use of *man* in the sample. Table 10 is from the study of Hall (2020, p. 122) and presents the findings in the whole MLE corpus.

Table 9. Man pronoun in the Grime corpus

	Frequency in the Grime corpus
1st singular	6
2nd singular	6
3rd singular	7
1st plural	0
2nd plural	3*
3rd plural	4*
impersonal	5
ambiguous	9

*= *man* modified by *them* or *you*

Table 10. Man pronoun in the MLE corpus

	Frequency in the MLE Corpus
1st singular	11
2nd singular	0
3rd singular	1
1st plural	1
2nd plural	0
3rd plural	0
impersonal	2
ambiguous	-

(Hall 2020, p. 122)

The results in table 10 are from the entire MLE corpus, so it is clear that the use of *man* as a pronoun was relatively rare in the MLE corpus. However, this does not mean that the use of *man* pronoun would be that rare among MLE speakers in general. The compilers of the MLE corpus, note that the *man* pronoun "is certainly more widespread than these figures suggest: it can be heard on the streets of London, […] and it can be heard on the media, both in spontaneous sustained discourse (for example in a YouTube Biography of the UK rapper Giggs 3) and in scripted UK television sitcoms set in multicultural areas of London" (Cheshire et al. 2013 p. 73). Only 6 speakers in the MLE corpus used *man* as a pronoun, so only data from those participants is included in the table. It should also be noted that table 10 only contains clear instances of *man* as a pronoun, as ambiguous cases were omitted in Hall's (2020) study.

In the MLE corpus, *man* was most often used either as its literal sense or as an address term. Cheshire (2015, p. 619) notes the widespread use of the word as an address term/discourse marker. Cheshire also notes the multifunctionality of *man*; it can be used to express surprise along with several other emotions, to add emphasis to a sentence, and "construct solidarity in discourse. Cheshire (2015, p. 625) claims that as an address term *man* has fully grammaticalized, and the number of results as an address term/discourse marker in the MLE corpus supports that.

The data from the Grime corpus was, however, much more fruitful. 40 out of the 200 instances of *man* in the corpus were instances where the word was used as a pronoun. This amount is staggering when compared to the MLE corpus. This would obviously suggest that the use of *man* as a pronoun would be especially common in the context of Grime. However, as noted in the previous paragraph, the MLE corpus does not offer a realistic view of the phenomena, so the two corpora cannot really be compared to one another.

Interpreting *man* was not an easy task, as the referent of the word, when used as a pronoun

is identified solely from the context (Hall 2020, p. 128). Sometimes it was not possible to identify which pronoun was *man* replacing from the context; thus there are 9 ambiguous cases in Table 9. Examples 20 and 21 showcase the problems with a lacking context, where the referent was unclear. Example 20 contains two instances of *man* used as pronoun. The second use in the tweet refers quite clearly to 1st person singular, but the first instance is unclear. The referent in the first context is the football club Liverpool F.C., but in terms of personal pronouns, both 1st and 3rd plural person readings are possible. It is unclear from this context alone to say whether Frisco includes himself as part of the team or is he referring to the team in the 3rd person sense.

(20) Whenever Liverpool lose, I get so many tweets from the hate brigade filled with laughing emojis. But when **man** get to the champions league final man can't even get a one congrats? You know who you are aswell 🙄🤔🌚🙂 (@BigFris, 2018-05-02, 991786601708769282)

(21) BIG E'S ON A BADMAN TING,**MAN** DONT GAS ME
😄😄😄😄😄😄😄😄💥💥💥💥💥💥💥💥💥 (@officialgiggs, 2017-11-08, 928141935134625792)

Cheshire (2015, p. 627) suggests that *man* is in the process of grammaticalizing as a first-person pronoun, and the findings from the MLE corpus would support that suggestion. However, the data of this study suggests otherwise. There were 6 clear instances of *man* being used as a first-person pronoun (although many ambiguous cases allowed a first person reading), with 2nd and 3rd person singulars being as common as the first-person usage. Example (22) showcases the use of *man* as a first-person pronoun in the Grime corpus.

(22) Big up my OG bredrin Alex from primary school. I went toilet at his yard, saw the bidet and was baffed for 20 minutes. Thought it was a water fountain. Lucky **man** weren't thirsty. RT @user @JmeBBK Bidet gang where you at? (@JmeBBK 2020-08-03, 1290380763880394757)

The second and third person uses of *man* as a pronoun were quite common in the sample. This might be affected by the conversational aspect of Twitter as many instances were found in the replies of the artists when they were conversing about other people. However, the MLE corpus also consisted mainly of interviews, so the conversational aspect was present in that data too. The use of *man* in the study of Hall (2020) as a second- or third-person pronoun was very rare, so it was surprising to see so many instances in the Grime corpus. Example (23) showcases the word being used as 2nd person singular pronoun and (24) as 3rd person singular.

(23) Man wanna act intelligent, I told my man he's irrelevant, what's **man** doing for the culture, what do dem yutes wanna be when their older […] (@jammerbbk, 2020-04-09, 1248323601016803328)

(24) @user @user last year man was out here hiring venues inviting anyone for free to come watch all type of MCs but he's clout chasing in 2020 causes he tweeted he wants to know who's about so he can roll with them too??? Does **man** know what his album name stands for? Or the live show concept??? (@KingPMoney, 2020-01-29, 1222552952038154245)

The results of 2nd and 3rd person plural cases are displayed with an asterisk on table 9 due to the fact that instances of *man* as those pronouns were modified by *you* in the case of 2nd person plural and *them* in 3rd person plural uses. These instances were left out in the study of Hall (2020), but they are included here, with due caution. *You man* and *them man* (as well as the 1st person plural *us man*) are regarded as noun uses of *man* in the studies of Cheshire (2015) and Hall (2020), but they often function as pronouns. Here, these forms are included in the table in order to showcase that *man* is used to replace plural person pronouns with the aforementioned modifications. *Man*, without these modifications, can be used in definite plural forms, as noted in Hall (2020, p. 127), but in the sample here, these forms were not found (although many impersonal readings and ambiguous cases allowed for definite plural readings). Example (25) showcases *man* as 2nd person plural and (26) as 3rd person plural pronoun. The 1st person plural was not found in the sample, but the whole corpus did include instances where *us man* was used in place of the 1st plural form *we* (Example 27)

(25) They flex with their cars and their chain but **them man** can't get yards in their name #Dc (@jammerbbk 2020-09-30, 1311273213746466817)

(26) Oiii **you Man** need to Get them bluku winter socks inn 😎https://t.co/hN28VtrZHN #goingfast (@DDoubleE7 2016-11-21, 800648438002176000)

(27) Blood look at my trainers **us man** are big inner di game!!!!! 😎😎😎 #blukumax1s https://t.co/x8oVmAA9aH (@DDoubleE7, 2016-01-03, 683616820604059648)

Impersonal pronoun here means the use of *man* to refer to an undefined group of people (the distinction of an impersonal pronoun is imitated from Hall [2020], see pages 123-127 for a wider discussion). 5 instances of impersonal uses were found in the sample. Examples (28) and (29) showcase the use of *man* as an impersonal pronoun. The use of the impersonal *man* was more common here than it was in Hall's (2020) study. However, it must be noted that the impersonal

readings of *man* often allow a definite personal interpretation as well (Hall 2020, p. 148) and some impersonal readings in the Grime corpus could have been definite personal uses as well given the right context.

(28) @user Trust me fam, I cant stand when **man** try bread it with me. I'm more Anti than my smile says. Just cool nuh. (@Novelist, 2020-06-27, 1276846053879435265)

(29) It's not about ends we should of left that in 2001 it's about unity and massive shows and murking... I hope that **man** understand that things everyone has had to deal with to make this shit safe and it still ain't 100% safe (@jammerbbk 2019-05-01, 1123547962716688385)

4.4.2 Gender differences

The results of this section are from the whole corpus, not from the sample that was used above. 214 instances of *man* were found in the female subcorpus. Here, the word was most often used as an address/reference term (Examples 30 and 31). In this case, *man* was gender-neutrally, as in many cases the term was used to address a female.

(30) @user Rest up and get well soon **man** x (@ladysov, 2016-08-04, 761266799707090944)

(31) @user That's that reallll grind🙏 just taking my nieces to school every blue morning is enough for me **man** 😂❤️(@IAMSHYSTIE, 2019-09-03, 1168799460115304448)

However, the use of *man* as a pronoun by females was rare in the corpus. Only 6 instances were recorded in the data here. Hall (2020, p. 127) and Cheshire et al. (2013, p. 624) note that the pronoun is also used by females, although the use seems to be rather rare, according to her. Cheshire (2015, p. 625) suggests that females might be avoid the pronoun because it indexes a masculine gender. This seems like a peculiar explanation since all females in this study did use *man* as an address term, where it could be thought to index masculinity as well. However, Cheshire (2015, p. 625) notes that as an address term/pragmatic marker *man* has already grammaticalized, whereas with the pronoun, the process is still in progress.

It should also be noted that the sample was rather small, which might have affected results here. However, according to this study and from results of previous researchers, it seems that the use of *man* as a pronoun is heavily linked with male users. Having said that, the use was still much more common in the Grime corpus than it has been in previous studies (Hall 2020; Cheshire 2015). Three out of the five females in the corpus did use the word as a pronoun and the instances can be found in the examples 32-37 below.

(32) But you can't just diss and come tell **man** sorry. (@ItsCleopatra 2016-04-29, 725943414890483712)

(33) @user @user Looool up north **man** loveeeeeee ah baseline omg😂.. 1min 22secs is a cold gif wait to happen, I need to grab that😂 (@IAMSHYSTIE 2020-06-18, 1273562879283941377)

(34) Almost every interview Ruiz did he's hailing up God. **Man** was moving with the higher powers and energies from this. He kept saying I'm gona cause the upset (@IAMSHYSTIE 2019-06-02, 1135064709147836417)

(35 "@user @Ms_Dynamite they try take **man** for a holiday maker like I am gonna be here 7 hours before a flight" >> 😂😂😂😩😩😩😭😭😭🙈🙈🙈😘😘😘 (@Ms_Dynamite 2014-07-27, 493461842006528001)

(36) @user **Mans** on the hunt.... Guess how much on eBay […] (@Ms_Dynamite 2014-12-08, 541922669709762560)

(37) "@jammerbbk: YOU KNOW WHAT **MAN** CANT EVEN MAKE ME VEX AGAIN ILL JUST WALK http://t.co/CLhppRP5QA">>LOVEEE THIS PIC!!! 😂😂😭😭😍😍😘😘.. Love bro x (@Ms_Dynamite 2014-07-04, 485024041867427842)

5. Discussion

Some discussion of the results was already present in the previous section 4. That analysis is continued and extended in this section. This section maps the main findings of the study and discusses them in relation to the research questions. As a reminder, the research questions are:

1. Are MLE elements used in written form?

2. Which features have made their way into written language?

3. Does the usage of elements differ between spoken and written mediums?

5.1 Discussion of phonological elements

The keyword analysis provided many phonological elements that had made their way into written language as well. Section 4.2 showcased the results in terms of how dental fricatives were realized by MLE users in written form. Dh-stopping was the most common phenomenon by far in terms of phonological elements, with others being much rarer. This section studies the phenomena presented in section 4.2 and discusses the phonological elements of the Grime corpus on a broader level.

As noted in section 2.6, phonological elements are used with many non-standard varieties in written language. Eisenstein (2015, p. 162) notes that the use of phonological spellings is often used to underline the fact that the user chooses to use a non-standard variety, and these choices "can index minority and subcultural identities". It must also be considered that the use of phonological elements is much more of a conscious choice in written form than it is in casual conversation. Thus,

it seems that the grime MCs in our study want to showcase a particular identity when using these elements. Drummond (2018, p. 192) argues that the use of Th-stopping in his study indexes street/grime values. The identification of indexical values of using phonological elements would require extensive qualitative analysis, and thus it is not feasible for the scope of this study. However, Drummond's argument above seems to make sense in this case as well; MCs want to emphasize their involvement in the grime scene by these linguistic choices since the context of these non-standard spellings would support that notion.

Contrary to popular belief, MLE speakers are very much able to switch between language variants depending on the context (Kircher & Fox 2019b, p. 4-5). The importance of context was mentioned with the spelling *yout* for *youth*, where it was clear that the standard spelling was used in more official context where the artists wanted to perhaps discard the identity of a Grime MC, discussed more serious matters, or were trying to reach a larger audience than the grime fanbase. This notion was evident in many other cases as well; oftentimes, phonological elements appeared in clusters or in contexts where other MLE features were also visible (Examples 38, 39 and 40). Examples 41 and 42 present instances where the artists chose to use standard language. Standard spellings seemed to appear in formal situations in the case of *youth/yout* but standard spellings were used in informal settings as well, and overall they were used in more widespread contexts than the non-standard ones. The context-dependency was more evident with word pairs, where the non-standard spelling appeared with a similar frequency to the standard one such as *youth/yout*, *thing/ting* etc.

(38) ALL THIS AFROBEAT VS DANCEHALL CHAT... BUT WHY NOT BOTH? LOL. BUT HERE'S SOME BASHMENT VIBES FI DI GYAL DEM. 'DJ FRASS FEAT. CHIP - LOCK DOWN' […] (@OfficialChip, 2020-05-08, 1258746652275458049)

(39) MAN CANT CHAT TO ME WHEN ITS A VIBEZ TING HAHAHAHAHAAHAHAHHAHAHAHAHHAH (@officialgiggs, 2017-12-02, 937032604431339520)

(40) J2k is a bloodclaat demon wid da flow Footsie no favours album out now! (@BigFris, 2020-06-06, 1269252468233420800)

(41) @user yeah you're right, I just feel sorry for them really, they've lost their child. Must be hell. @IamBabyBlue, 2019-03-16, 1106860177712664576)

(42) I Was listening to my first album today. It's mad how my mindset was back then. I've actually over achieved my objectives. Growth is a beautiful thing. (@LethalBizzle, 2021-02-02, 1356624905618546690)

One interesting aspect, that was already found in the analysis of keywords, was the fact that the word *brother* was a subject of both Dh-stopping and -fronting. *Bruv* and *bruva* were instances where Dh-fronting was used, whereas *brudda* and *bredda* represented Dh-stopped instances. JME, Frisco, Chip, Lethal Bizzle, Novelist, and President T used both stopped and fronted variants, while others such as Jammer, Devlin, and P Money only used the Th-fronted variants. On the other hand, Ghetts, Giggs, and D Double E only used the Dh-stopped variants, and they were responsible for 55 of the 99 instances of *brudda*/*bredda*. The three MCs all come from Caribbean backgrounds, which probably affect the results, as Dh-stopping is a prevalent feature of many Caribbean creoles. This notion shows the diversity of MLE's feature pool; different speakers utilize different elements in their language, and clearly this phenomenon is also visible in written form too.

The original idea of this study was to compare the usage of MLE elements in written form versus the spoken form. However, this was not possible with phonological elements due to the fact that the features that were examined were almost non-existent in the MLE corpus (*ting* appearing 4 times, of which 1 occurrence was relevant, *Dat* appearing twice, etc.) Perhaps these features genuinely did not appear in the participants' speech, but that seems unlikely since the examined features are noted as MLE features in many publications (e.g. Drummond 2018, Cheshire et al. 2008). Thus, we must question the transcription of the MLE corpus. The MLE corpus did include some words that can be considered phonological (e.g. *aks, cus, dunno, duh* etc.). This raises the question whether the transcription of the corpus is partially phonological and partially not or were the phonological elements regarding dental fricatives genuinely that rare in the corpus.

To sum up, phonological elements were found in the Grime corpus, but the number of significant findings was limited to a handful of words. *Da, dat, dem,* and *ting* were used widely across the participants with, other words such as *brudda, yout, dis* and *di* also being used by at least half the artists. However, it must be questioned whether some of these words are even phonological innovations. Eisenstein (2013, p. 11) notes that "Even if social media authors introduce new orthographic transcriptions to capture the sound of language in the dialect that they speak, such innovations may be purely lexical". This might be the case with some of the items that were researched in the study. However, further research would be needed in order to seek out whether these orthographic choices are lexical or phonological.

5.2 Discussion of slang words

Section 4.3 showcased that the use of slang words is a feature that has clearly made its way to written form as well. Table 5 displayed the results in terms of slang words, and it was clear that MLE users (at least grime artists) use slang terms frequently in written form. This finding was not that surprising because slang terms are widely used on Twitter and social media in general.

The original idea was to mainly use slang words from the MLE corpus in order to compare the two datasets. Keywords from the MLE Corpus were collected to find MLE slang words. However, the keyword analysis did not offer many slang words (Although this was also affected by the fact that address terms were omitted from analysis). Thus, external sources, such as previous research papers and news articles, and most importantly, the MLE glossary by Nott (2013), were also used to collect the slang words. This was done to find more MLE elements and get a fuller picture of the usage of slang words.

Here again, the MLE corpus showed its weaknesses as MLE slang words were relatively scarce in the corpus. Unfortunately, the two corpora differ from one another quite substantially, so the comparison of frequencies was not very relevant. Instances of words that were more common in the MLE corpus were actually really rare. Thus, it would seem that the usage of slang terms would be more common in written form, but a more logical explanation would be that the context of grime supports the use of slang terminology. As the comparison between the two datasets is not applicable, reliable conclusions cannot be made on the basis of this data.

The notion about the artists switching registers, presented in the previous section 5.1, was evident within slang words as well. In many cases, slang words appeared in sentences/contexts that included other non-standard elements too. Standard English spellings, lexical choices etc. appeared in the corpus, but they were often used when conversing about more serious matters or when they were communicating to a wider audience to that of grime. This supports the notion presented in 5.1 that, MLE speakers do switch registers in different contexts.

As noted, the list of slang words presented in 4.3 is not comprehensive. Several slang words, such as *peak, safe, bare, bait, whip, air,* and *long*, were omitted because their non-slang meanings frequently appeared in the corpora. The number of slang words was sufficient for the purposes of this study even without the aforementioned terms, and thus their exclusion was not too significant considering the overall results.

5.3 Discussion of *man*

The results regarding the *man* pronoun were interesting, when compared to previous research. In Cheshire's study (2015, p. 614) 66 out of the 94 instances of *man* as a pronoun were found in 1st

person singular positions, 13 at 1st plural, with other positions quite rare. However, the results of this study contradict many of the findings of that study. The grammaticalization of *man* as a first-person pronoun does not seem to be happening in written form; all singular forms, as well as the impersonal use of *man* were as common as the first-person usage in the sample. The context of grime does not seem to affect this either, since one of Cheshire's (2015) data sources was the *Giggs Biography*, which features grime artists and fans. In *Giggs Biography*, the results heavily favored the use of *man* as a 1st person pronoun (9 out of 11 tokens being 1st person uses). It is hard to draw any reliable conclusions from the relatively small sample here, but it seems that Cheshire's hypothesis of *man* grammaticalizing as a first-person pronoun would not be correct at least in the context of this study.

The results of *Man* pronoun in plural positions were also fascinating. There were no clear instances where *man* would take the place of a plural pronoun, but the noun *man* was used with premodifications *us* (1st plural), *you* (2nd plural) and *them* (3rd plural) with pronominal values. These forms were found often in the corpus to replace the standard pronouns *we, you* and, *they*. This would indicate that perhaps MLE users prefer to use the aforementioned premodifications with *man* when they are replacing standard pronouns with the word *man*. However, it should also be noted that this phenomenon is briefly mentioned in the studies of Hall (2020) and Cheshire (2015). *Us man*, *you man*, and *them man* are noun uses of *man* but the fact that they can function as pronouns in a sentence is certainly a fascinating trend.

The grammaticalization process of *man* as a pronoun is still in progress, as females used *man* widely as a reference/address term, but instances of the pronoun use were relatively rare. However, the use was more common in the Grime Corpus than it has been on previous research papers (Hall 2020; Cheshire 2015). (Hall 2020, p. 127) suggests that *man* would not be restricted to males as referents but does not have examples of the use of *man* referring to females. However, examples 35 and 36 of this study show Ms. Dynamite using *man* as a first-person pronoun, and there were examples in the corpus that revealed that both men and women used *man* as a pronoun to refer to females in the corpus.

The sheer number of results from the Grime Corpus was staggering compared to previous studies (especially when compared to the MLE corpus). However, the results of this study cannot be regarded conclusive. The sample of this study is rather small. Moreover, the context of grime could affect the results too, and written context might also be a factor for the differing results between this study and the work of previous researchers. However, the findings regarding *man* are certainly fascinating, and could indicate that *man* as a pronoun is used differently than what previous

researchers have hypothesized. That said, more research is needed to investigate the properties of *man* further.

6. Conclusions

This study aimed to discover whether MLE is used in written form and how users utilize the different elements of MLE in online communication. From an empirical standpoint, I had already seen many MLE elements appearing online, so it was clear that the variety was also used in written form to some extent. However, this study provides evidence that features of the multiethnolect have spread to social media writing, at least by people in the grime scene. The inclusion of retweets allowed to get a wider scope of participants and proved that MLE elements are used beyond the artists included in this study. Some might argue that this study is about the language of grime, which is obviously true. However, as emphasized several times during this paper, the language of grime and MLE are very much connected to one another. This study sees the language of grime as part of MLE, existing as part of MLE's feature pool.

Phonological elements, slang words, and *man* pronoun were all visible in the data. Several other features of MLE that were not individually analyzed in this study (e.g. non-standard way of using past tense *be*, the quotative expression *be like*, and MLE intensifiers) were all found in the corpus. Thus, it is clear that MLE is used in online communications as well. However, the context of grime could be a factor that affected the results, and it would be fascinating to see whether MLE speakers outside of the grime scene would utilize MLE features differently in written form.

The use of phonological elements was present in the corpus, but it was limited to a handful of words. Dh-stopping was by far the most prevalent phonological feature. Th-stopping was limited to words *ting* and *yout*, and Th- and Dh-fronting were almost non-existent in the corpus. As a whole, the use of phonological elements was not widespread. The use of slang words, on the other hand, was very frequent in the Grime Corpus. Almost all slang words were more common in the Twitter data than in the MLE corpus. Similarly, the use of *man* as a pronoun was considerably more evident in the Grime Corpus. The results of *man* also suggest that the word would not be grammaticalizing as a first-person pronoun like Cheshire (2015, p. 627) had proposed. Instead, it was used evenly to replace all singular person pronouns. *Man* was also used to replace plural person pronouns; this was done with premodifications *us, you, them*.

There are still many areas of MLE that need to be further research. Adams' (2018, p. 13) highlights the value of using grime to research British multi-ethnic dialects. This study proves that the sociolinguistic landscape of grime is fertile in terms of MLE and urges future researchers of MLE to use the context of grime as a viewpoint to study this new language variety. The indexical

links of using MLE elements were hypothesized in parts of this study. However, that was not the focus of the thesis, but it could be an area for future researchers to investigate.

This study does have its weaknesses. The comparison between spoken and written mediums was not possible in many parts due to several reasons. Firstly, as noted in many instances during this paper, the MLE corpus was lacking in terms of phonological features and many other MLE elements. Secondly, the context of the two corpora differed vastly, and the context of grime is undoubtedly one thing that affected the results here. This study does not consider the diversity of MLE speakers adequately, as the participants are all grime artists. A more diverse group of participants would be needed to make reliable conclusions about MLE speakers writing conventions. It should also be noted that this study does not prove whether elements have spread outside the use of Twitter. Consequently, the differences between written and spoken mediums cannot be regarded as decisive. Thus, further research is needed in order to examine the possible differences between written and spoken MLE.

In conclusion, this study proves that MLE has spread to online communication, at least with grime artists on Twitter. As well as that, this study discovered possibly new innovations of the language variety (most notably the findings regarding *man*). The results are fascinating, but they should not be used to make any definite conclusions. Instead, this study hopes to act as a starting point for future studies to develop and expand on.

References

Adams, R., 2019. “Home sweet home, that’s where I come from, where I got my knowledge of the road and the flow from”: Grime music as an expression of identity in postcolonial London. Popular Music & Society, [online] 42(4), pp.438–455. Available at: <https://doi.org/10.1080/03007766.2018.1471774 > [Accessed 22 Jan. 2021].

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, [online] 60, pp.11–27. Available at: < https://doi.org/10.1016/j.langcom.2018.01.004> [Accessed 21 Jan. 2021].

Ahmed, W., Bath, P. and Demartini, G., 2017. Chapter 4: Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges. pp.79–107.

BBC 2018. *BBC Radio 4 - Radio 4 in Four - 17 Multicultural London English words and what they mean*. [online] Available at: <DOI: 10.1108/S2398-601820180000002004> [Accessed 23 Apr. 2021].

Bolinger, D. 1972. Degree Words. The Hague, Paris: Mouton de Gruyter.

Boyd, D., Golder, S. and Lotan, G., 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In: 2010 43rd Hawaii International Conference on System Sciences. 2010 43rd Hawaii International Conference on System Sciences. pp.1–10. <DOI: 10.1109/HICSS.2010.412>

Blodgett, S.L., Green, L. and O’Connor, B., 2016. Demographic Dialectal Variation in Social Media: A Case Study of African-American English. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. [online] EMNLP 2016. Austin, Texas:

Association for Computational Linguistics.pp.1119–1130. Available at: < DOI:10.18653/v1/D16-1120 > [Accessed 14 Apr. 2021].

Cambridge Dictionary. Available at: <https://dictionary.cambridge.org/> [Accessed 14 Apr. 2021]

Cheshire, J., 2013. Grammaticalisation in social context: The emergence of a new English pronoun. Journal of Sociolinguistics, 17. <DOI: 10.1111/josl.12053>

Cheshire, J., 2020. Taking the longer view: Explaining Multicultural London English and Multicultural Paris French. Journal of Sociolinguistics, [online] 24(3), pp.308–327. Available at: < https://doi.org/10.1111/josl.12385> [Accessed 7 Jan. 2021].

Cheshire, J., Fox, S., Kerswill, P. and Torgersen, E., 2008. Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. Sociolinguistica, 22, pp.1–23. <DOI: 10.1515/9783484605299.1>

Cheshire, J., Kerswill, P., Fox, S. and Torgersen, E., 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. Journal of Sociolinguistics, [online] 15(2), pp.151–196. Available at: <DOI: 10.1515/9783484605299.1> [Accessed 14 Oct. 2020].

Cheshire et al., 2011. Multicultural London English Corpus. Available at: < https://www.sketchengine.eu/london-english-corpus/> [Accessed 1 Nov. 2020].

Cheshire, J., Fox, S., Kerswill, P. and Torgersen, E., 2013. Language contact and language change in the multicultural metropolis. Revue francaise de linguistique appliquee, [online] Vol. XVIII(2), pp.63–76. Available at: <https://www.cairn-int.info/article.php?ID_ARTICLE=E_RFLA_182_0063> [Accessed 7 Jan. 2021].

Cheshire, J., Nortier, J. and Adger, D., 2015. Emerging multiethnolects in Europe. [online] undefined. Available at: </paper/Emerging-multiethnolects-in-Europe-Cheshire-Nortier/6133a18dacbbaff0e68c7ab87b269e54b2600f94> [Accessed 3 Jan. 2021].

Clyne, M., 2000. Lingua Franca and Ethnolects in Europe and Beyond. <DOI: https://doi.org/10.1515/9783110245196.83>

Davila, B., 2016. The Inevitability of "Standard" English: Discursive Constructions of Standard Language Ideologies. *Written Communication*, [online] 33(2), pp.127–148. Available at: <https://doi.org/10.1177/0741088316632186> [Accessed 28 Apr. 2021].

Dedman, T., 2011. Agency in UK hip-hop and grime youth subcultures - peripherals and purists. Journal of Youth Studies, [online] 14(5), pp.507–522. Available at: <DOI: 10.1080/13676261.2010.549820> [Accessed 22 Jan. 2021].

Drummond, R., 2017. (Mis)interpreting urban youth language: white kids sounding black? Journal of Youth Studies, [online] 20(5), pp.640–660. Available at: <DOI: 10.1080/13676261.2016.1260692> [Accessed 23 Oct. 2020].

Drummond, R., 2018. Maybe it's a grime [t]ing: th-stopping among urban British youth. Language in Society, [online] 47(2), pp.171–196. Available at: <DOI: 10.1017/S0047404517000999> [Accessed 11 Jan. 2021].

Eisenstein, J., 2013. Phonological Factors in Social Media Writing. In: Proceedings of the Workshop on Language Analysis in Social Media. [online] Atlanta, Georgia: Association for Computational Linguistics.pp.11–19. Available at: <https://www.aclweb.org/anthology/W13-1102> [Accessed 16 Apr. 2021].

Eisenstein, J., 2015. Systematic patterning in phonologically-motivated orthographic variation. Journal of Sociolinguistics, [online] 19(2), pp.161–188. Available at: <https://doi.org/10.1111/josl.12119> [Accessed 16 Apr. 2021].

Goldbeck, J., 2018. What is MLE, who speaks it, and is it safe? Scripps Senior Theses. [online] Available at: <https://scholarship.claremont.edu/scripps_theses/1172>.

Green 2014. Multicultural London English: the new 'youthspeak'. [online] Routledge. Available at: <DOI: 10.4324/9781315857787-13> [Accessed 14 Apr. 2021].

Hall, D., 2020. The impersonal gets personal. Natural Language & Linguistic Theory, [online] 38(1), pp.117–150. Available at: <DOI: 10.1007/s11049-019-09447-w> [Accessed 14 Oct. 2020].

Holden, S. 2017. Academic study finds grime as 'disruptive and powerful' as punk. *BBC News*. [online] 11 Oct. Available at: <https://www.bbc.com/news/newsbeat-41559356> [Accessed 3 May 2021].

Jones, T., 2015. Toward a Description of African American Vernacular English Dialect Regions Using "Black Twitter". American Speech, [online] 90(4), pp.403–440. Available at: <https://doi.org/10.1215/00031283-3442117> [Accessed 25 Jan. 2021].

Jørgensen, A., Hovy, D. and Søgaard, A., 2015. Challenges of studying and processing dialects in social media. In: Proceedings of the Workshop on Noisy User-generated Text. [online] Beijing, China: Association for Computational Linguistics.pp.9–18. Available at: <https://www.aclweb.org/anthology/W15-4302> [Accessed 25 Jan. 2021].

Kerswill, P; Cheshire, J; Fox, S; & Torgersen, E. 2007. Linguistic Innovators: the English of Adolescents in London. Final Report of ESRC funded project: RES-000-23-0680. Swindon: ESRC.

Kerswill, P., 2013a. Identity, ethnicity and place: the construction of youth language in London. In: P. Auer, M. Hilpert, A. Stukenbrock and B. Szmrecsanyi, eds. *Space in language and linguistics*. [online] Walter de Gruyter.pp.128–164. Available at: <http://eprints.whiterose.ac.uk/75329/> [Accessed 23 Apr. 2021].

Kerswill, P., 2013b. Multicultural London English: the Rise of a New Phonological Variety? University of York. Lecture given to students and staff at the Department of English, University of Zurich.

Kerswill, P., 2014. The objectification of 'Jafaican' : the discoursal embedding of Multicultural London English in the British media. Available at: <DOI: 10.1515/9783110346831.427> [Accessed 10 Jan. 2021]

Kircher, R. & Fox, S., 2019a. Attitudes towards Multicultural London English: implications for attitude theory and language planning, Journal of Multilingual and Multicultural Development, 40:10, 847-864. Available at:<DOI: 10.1080/01434632.2019.1577869> [Accessed 19 Jan. 2021]

Kircher, R. & Fox, S., 2019b. Multicultural London English and its speakers: a corpus-informed discourse study of standard language ideology and social stereotypes, Journal of Multilingual and Multicultural Development. Available at: <DOI: 10.1080/01434632.2019.1666856> [Accessed 20 Jan. 2021]

Kohonen, I; Kuula-Luumi, A; & Spoof, S. 2019. The ethical principles of research with human participants and ethical review in the human sciences in Finland. Available at: < https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf > [Accessed 15 Apr. 2021].

Lippi-Green, R. 2012. English with an Accent: Language, Ideology, and Discrimination in the United States. 2nd ed. London: Routledge.

Luger, E., Moran, S. and Rodden, T., 2013. Consent for all: revealing the hidden complexity of terms and conditions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13. [online] New York, NY, USA: Association for Computing Machinery.pp.2687–2696. Available at: <https://doi.org/10.1145/2470654.2481371> [Accessed 24 Feb. 2021].

Macmillan Dictionary. Available at: <https://www.macmillandictionary.com/> [Accessed 19 Apr. 2021]

Manner, M; Teittinen, P. 2021. *Nuoret | HS seurasi koko talven nuorten elämää Helsingin asemilla ja ostoskeskuksissa – Avautui todellisuus, josta aikuiset eivät tiedä juuri mitään*. [online] Helsingin Sanomat. Available at: <https://www.hs.fi/sunnuntai/art-2000007868967.html> [Accessed 18 Apr. 2021].

Marcus, M.P., Marcinkiewicz, M.A. and Santorini, B., 1993. Building a large annotated corpus of English: the penn treebank. Computational Linguistics, 19(2), pp.313–330.

Martínez, I.P, 2013. Non-standard negation in modern English: a corpus-based study of four salient features. [online] undefined. Available at: </paper/Non-standard-negation-in-modern-English%3A-a-study-of-Mart%C3%ADnez/7215ffd77d0b706dc6b08f95af28f57ec4d2df2b> [Accessed 24 Feb. 2021].

Martínez, I.P., 2015. Variation, development and pragmatic uses of innit in the language of British adults and teenagers1. English Language & Linguistics, [online] 19(3), pp.383–405. Available at: <DOI: 10.1017/S1360674314000288> [Accessed 13 Jan. 2021].

Martínez, I.P. and Pertejo, P.N., 2015. "Go up to miss thingy". "He's probably like a whatsit or something".: Placeholders in focus. The differences in use between teenagers and adults in spoken

English. Pragmatics, [online] 25(3), pp.425–451. Available at: <DOI: 10.1075/prag.25.3.05pal> [Accessed 24 Feb. 2021].

Marwick, A. and Boyd, D., 2011. To See and Be Seen: Celebrity Practice on Twitter. Convergence, [online] 17(2), pp.139–158. Available at: <https://doi.org/10.1177/1354856510394539> [Accessed 12 Feb. 2021].

Milroy, J. 2001. "Language Ideologies and the Consequences of Standardization." Journal of Sociolinguistics 5 (4): 530–555

Nguyen, D.-P., 2019. *Dialect Variation on Social Media*. [online] Available at: </paper/Dialect-Variation-on-Social-Media-Nguyen/78ec4158efd8f031176feaf7f57263d00d9d4082> [Accessed 19 Apr. 2021].

nikoma, 2015. List of female grime MCs. [Reddit Post] r/grime. Available at: <www.reddit.com/r/grime/comments/31vvyj/list_of_female_grime_mcs/> [Accessed 12 Feb. 2021].

Nott, C. 2013. Multiethnic London English Glossary. Available at: <https://language-and-innovation.com/2016/10/24/multethnic-london-english-a-glossary/> [Accessed 20 Apr. 2021].

Núñez-Pertejo, P. and Palacios-Martínez, I., 2018. Intensifiers in MLE: New trends and developments1. Nordic Journal of English Studies, [online] 17(2), pp.116–155. Available at: <http://njes-journal.com/articles/abstract/10.35360/njes.436/> [Accessed 24 Feb. 2021].

Page, R., Barton, D., Unger, J. and Zappavigna, M., 2014. Researching language and social media: A student guide. Available at: <DOI: 10.4324/9781315771786>

Palacios Martínez, Ignacio M. 2017. The Emergence of a New Variety of English: The Multicultural London English. Nexus, AEDEAN: 37 - 53.

plain_lou, 2019. THE GRIME ESSENTIAL LIST (UPDATE) [2019 EDITION]. [Reddit Post] r/grime. Available at: <www.reddit.com/r/grime/comments/c8nmqn/the_grime_essential_list_update_2019_edition/> [Accessed 12 Feb. 2021].

Quist, P., 2008. Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. International Journal of Bilingualism, [online] 12(1–2), pp.43–61. Available at: <https://doi.org/10.1177/13670069080120010401> [Accessed 4 Jan. 2021].

Rampton, Ben. 2015. Contemporary urban vernaculars. In Jacomine Nortier and Bente A. Svendsen (eds.) Language, Youth and Identity in the 21st Century: Linguistic Practices across Urban Spaces. 24-44, Cambridge University Press.

Schleef, E & Ramsammy, M 2013, 'Labiodental fronting of // in London and Edinburgh: a cross-dialectal study', English Language and Linguistics, vol. 17, no. 01, pp. 25-54. Available at: <https://doi.org/10.1017/S1360674312000317>

Seargeant, P. and Tagg, C. eds., 2014. The Language of Social Media: Identity and Community on the Internet. [online] Palgrave Macmillan UK. Available at: <DOI: 10.1057/9781137029317> [Accessed 1 Feb. 2021].

Sebba, M. 1993. London Jamaican. London, Longman

SketchEngine 2015a. English Penn Treebank POS tagset | Sketch Engine. Available at: <https://www.sketchengine.eu/penn-treebank-tagset/> [Accessed 6 Apr. 2021].

SketchEngine 2015b. Simple maths with keywords and terms | Sketch Engine. Available at: <https://www.sketchengine.eu/documentation/simple-maths/> [Accessed 7 Apr. 2021].

SketchEngine 2021a. Dashboard | Sketch Engine. [online] Available at: <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Flecorpus_3> [Accessed 12 Feb. 2021].

SketchEngine 2021b. enTenTen — English corpus from the web | Sketch Engine. Available at: <https://www.sketchengine.eu/ententen-english-corpus/> [Accessed 6 Apr. 2021].

Torgersen, E.N., Gabrielatos, C., Hoffmann, S. and Fox, S., 2011. A corpus-based study of pragmatic markers in London English. [online] 7(1), pp.93–118. Available at: <DOI: 10.1515/cllt.2011.005> [Accessed 22 Feb. 2021].

The Economist 2021. Grime and UK drill are exporting multicultural London English. *The Economist*. [online] 30 Jan. Available at: <https://www.economist.com/britain/2021/01/30/grime-and-uk-drill-are-exporting-multicultural-london-english?fbclid=IwAR3lLg0C8BuO19Du7mgWADuyh_cCcJEDrWEUFw7AUTY8OdMldSDgTdhUDtY> [Accessed 18 Apr. 2021].

Twitter 2021. Privacy Policy. [online] Available at: <https://twitter.com/en/privacy> [Accessed 23 Feb. 2021].

Urban Dictionary. Available at: <https://www.macmillandictionary.com/> [Accessed 14 Feb. 2021]

Vicinitas 2021. Vicinitas : Twitter Analytics Tool for Tracking Hashtags, Keywords, and Accounts. [online] Available at: <https://www.vicinitas.io/> [Accessed 12 Feb. 2021].

Zappavigna, M., 2011. Ambient affiliation: A linguistic perspective on Twitter. New Media & Society, [online] 13(5), pp.788–806. Available at: <https://doi.org/10.1177/1461444810385097> [Accessed 20 Apr. 2021].

Appendices

Appendix 1. Keywords of the whole Grime corpus (Reference corpus: EnTenTen)

Item	Frequency (focus)	Frequency (reference)	Relative frequency (focus)	Relative frequency (reference)	Score
"rt"	17508	185342	19438.60000	7.20000	2377.400
"skepta"	449	1261	498.50000	0.00000	476.300
"grime"	882	29557	979.30000	1.10000	457.100
"ghetts"	338	112	375.30000	0.00000	374.600
"frisco"	393	16849	436.30000	0.70000	264.700
"loool"	230	250	255.40000	0.00000	253.900
"giggs"	268	5411	297.60000	0.20000	246.800
"kano"	373	18176	414.10000	0.70000	243.700
"bro"	990	91441	1099.20000	3.50000	242.300
"dizzee"	227	1362	252.00000	0.10000	240.400
"jme"	221	1484	245.40000	0.10000	233.000
"banger"	333	15578	369.70000	0.60000	231.200
"ting"	517	40896	574.00000	1.60000	222.600
"fam"	404	32492	448.50000	1.30000	199.100
"grm"	195	2634	216.50000	0.10000	197.400
"hoodies"	173	1083	192.10000	0.00000	185.300
"looool"	159	155	176.50000	0.00000	176.500
"bruv"	150	454	166.50000	0.00000	164.600
"riddim"	155	1761	172.10000	0.10000	162.000
"ghett"	134	35	148.80000	0.00000	149.600
"kmt"	181	9989	201.00000	0.40000	145.600
"loooool"	127	89	141.00000	0.00000	141.500
"lol"	2287	451579	2539.20000	17.50000	137.400
"lool"	123	489	136.60000	0.00000	135.000
"icymi"	124	1410	137.70000	0.10000	131.500
"newham"	147	7924	163.20000	0.30000	125.700
"nah"	286	40943	317.50000	1.60000	123.200
"looooool"	107	56	118.80000	0.00000	119.500
"brudda"	107	153	118.80000	0.00000	119.100
"lotm"	110	940	122.10000	0.00000	118.800
"insta"	126	5757	139.90000	0.20000	115.200
"rascal"	189	22593	209.80000	0.90000	112.500
"coz"	177	20627	196.50000	0.80000	109.800
"cleo"	145	13252	161.00000	0.50000	107.100
"defo"	104	2371	115.50000	0.10000	106.700
"bizzle"	97	497	107.70000	0.00000	106.600
"oi"	194	28558	215.40000	1.10000	102.800
"vibes"	219	37498	243.10000	1.50000	99.600

"bratt"	95	3227	105.50000	0.10000	94.600
"mixtape"	130	14813	144.30000	0.60000	92.400
"jammer"	131	16075	145.40000	0.60000	90.300
"jackuum"	79	3	87.70000	0.00000	88.700
"ffs"	113	12949	125.50000	0.50000	84.200
"loooooool"	72	24	79.90000	0.00000	80.900
"bluku"	71	5	78.80000	0.00000	79.800
"nutcrackerz"	69	2	76.60000	0.00000	77.600
"hahaha"	129	22470	143.20000	0.90000	77.100
"remix"	223	59506	247.60000	2.30000	75.200
"stormzy"	72	2198	79.90000	0.10000	74.600
"tbh"	112	17581	124.40000	0.70000	74.600
"mz"	92	10756	102.10000	0.40000	72.800
"haha"	306	95689	339.70000	3.70000	72.400
"luv"	111	18620	123.20000	0.70000	72.200
"loll"	79	6742	87.70000	0.30000	70.300
"dem"	253	82294	280.90000	3.20000	67.300
"fuckin"	122	26613	135.50000	1.00000	67.200
"mandem"	60	160	66.60000	0.00000	67.200
"thanku"	60	697	66.60000	0.00000	65.800
"ppl"	157	45559	174.30000	1.80000	63.400
"ur"	299	109941	332.00000	4.30000	63.300
"jools"	63	3206	69.90000	0.10000	63.100
"tbf"	62	2914	68.80000	0.10000	62.800
"looooooool"	55	17	61.10000	0.00000	62.000
"tryna"	71	7514	78.80000	0.30000	61.800
"prez"	69	6926	76.60000	0.30000	61.200
"freestyle"	183	60514	203.20000	2.30000	61.100
"bday"	66	5794	73.30000	0.20000	60.700
"ep"	441	185273	489.60000	7.20000	60.000
"jheeze"	52	13	57.70000	0.00000	58.700
"feat"	398	173487	441.90000	6.70000	57.400
"adz"	52	740	57.70000	0.00000	57.100
"prod"	150	50418	166.50000	2.00000	56.800
"mc"	357	159389	396.40000	6.20000	55.400
"sparko"	49	93	54.40000	0.00000	55.200
"congrats"	174	65520	193.20000	2.50000	54.900
"bbk"	52	1982	57.70000	0.10000	54.500
"duppy"	49	674	54.40000	0.00000	54.000
"fav"	90	22821	99.90000	0.90000	53.600
"tonight"	1092	565574	1212.40000	21.90000	53.000
"nizzle"	46	113	51.10000	0.00000	51.800
"hun"	105	33620	116.60000	1.30000	51.100

"omg"	170	70371	188.70000	2.70000	50.900
"raskit"	45	19	50.00000	0.00000	50.900
"ft"	665	354039	738.30000	13.70000	50.300
"wid"	63	10896	69.90000	0.40000	49.900
"looooooooool"	44	13	48.90000	0.00000	49.800
"ima"	74	18704	82.20000	0.70000	48.200
"yeh"	81	22948	89.90000	0.90000	48.100
"naa"	65	13851	72.20000	0.50000	47.600
"ya"	485	267227	538.50000	10.30000	47.500
"hahahaha"	53	6842	58.80000	0.30000	47.300
"gonna"	750	430275	832.70000	16.70000	47.200
"rah"	54	7549	60.00000	0.30000	47.200
"smh"	64	13805	71.10000	0.50000	47.000
"wanna"	351	189372	389.70000	7.30000	46.900
"pre-order"	120	48826	133.20000	1.90000	46.400
"footsie"	44	2099	48.90000	0.10000	46.100
"rudeboy"	41	257	45.50000	0.00000	46.100
"gunna"	47	4022	52.20000	0.20000	46.000
"wtf"	110	43872	122.10000	1.70000	45.600
"sn1"	42	1207	46.60000	0.00000	45.500
"thankyou"	83	27383	92.20000	1.10000	45.200
"nang"	58	12129	64.40000	0.50000	44.500
"nuh"	51	7921	56.60000	0.30000	44.100
"dat"	134	62195	148.80000	2.40000	43.900
"skint"	41	1808	45.50000	0.10000	43.500
"badman"	41	1980	45.50000	0.10000	43.200
"tho"	206	111768	228.70000	4.30000	43.100
"vid"	97	39671	107.70000	1.50000	42.900
"mad"	663	420404	736.10000	16.30000	42.700
"init"	80	28701	88.80000	1.10000	42.500
"deffo"	39	1102	43.30000	0.00000	42.500
"lockdown"	63	17299	69.90000	0.70000	42.500
"retweet"	59	14906	65.50000	0.60000	42.200
"bruva"	37	35	41.10000	0.00000	42.000
"xmas"	80	29506	88.80000	1.10000	41.900
"pls"	59	15753	65.50000	0.60000	41.300
"hahahahaha"	40	2617	44.40000	0.10000	41.200
"hahah"	40	3042	44.40000	0.10000	40.600
"facking"	36	271	40.00000	0.00000	40.500
"xx"	170	95243	188.70000	3.70000	40.500
"devlin"	57	15821	63.30000	0.60000	39.900
"merch"	47	9069	52.20000	0.40000	39.400
"spotify"	122	64517	135.50000	2.50000	39.000

"ic3"	37	2425	41.10000	0.10000	38.500
"gigg"	35	1065	38.90000	0.00000	38.300
"tix"	40	5013	44.40000	0.20000	38.000
"popcaan"	34	506	37.70000	0.00000	38.000
"yo"	163	99066	181.00000	3.80000	37.600
"mwah"	34	926	37.70000	0.00000	37.400
"rapper"	151	91670	167.70000	3.50000	37.100
"mazza"	36	2971	40.00000	0.10000	36.700
"album"	2003	1542218	2223.90000	59.70000	36.600
"wamp"	34	1643	37.70000	0.10000	36.400
"aint"	78	36349	86.60000	1.40000	36.400
"lolll"	32	112	35.50000	0.00000	36.400
"gyal"	34	1748	37.70000	0.10000	36.300
"innit"	35	2579	38.90000	0.10000	36.200
"lmao"	53	16922	58.80000	0.70000	36.200
"nigga"	96	51252	106.60000	2.00000	36.000
"hahahah"	33	1291	36.60000	0.00000	35.800
"gotta"	255	179204	283.10000	6.90000	35.800
"dawg"	44	10321	48.90000	0.40000	35.600
"itscleopatra"	31	0	34.40000	0.00000	35.400
"dunno"	86	45062	95.50000	1.70000	35.200
"woah"	42	9407	46.60000	0.40000	34.900
"u"	1935	1570125	2148.40000	60.80000	34.800
"collab"	41	8725	45.50000	0.30000	34.800
"yea"	151	100295	167.70000	3.90000	34.500
"headie"	30	77	33.30000	0.00000	34.200
"ten10"	30	101	33.30000	0.00000	34.200
"sbtv"	30	251	33.30000	0.00000	34.000
"holla"	34	3804	37.70000	0.10000	33.800
"lmfao"	33	3283	36.60000	0.10000	33.400
"pan-fried"	34	4142	37.70000	0.20000	33.400
"shit"	534	433754	592.90000	16.80000	33.400
"mate"	533	433301	591.80000	16.80000	33.300
"babe"	193	141205	214.30000	5.50000	33.300
"truss"	87	51070	96.60000	2.00000	32.800
"dere"	35	5586	38.90000	0.20000	32.800
"yessss"	30	1213	33.30000	0.00000	32.800
"fris"	29	410	32.20000	0.00000	32.700
"cmon"	32	3416	35.50000	0.10000	32.300
"aswell"	50	19871	55.50000	0.80000	31.900
"wiley"	127	89413	141.00000	3.50000	31.800
"ldn"	35	6682	38.90000	0.30000	31.700
"kojo"	30	2166	33.30000	0.10000	31.700

"da"	405	342406	449.70000	13.30000	31.600
"jollof"	28	625	31.10000	0.00000	31.300
"hus"	39	11014	43.30000	0.40000	31.100
"yizzy"	27	6	30.00000	0.00000	31.000
"looooooooool"	27	10	30.00000	0.00000	31.000
"vibez"	27	154	30.00000	0.00000	30.800
"dubplate"	27	370	30.00000	0.00000	30.500
"buss"	56	28593	62.20000	1.10000	30.000
"mics"	33	6600	36.60000	0.30000	30.000
"jheeeze"	26	1	28.90000	0.00000	29.900
"yute"	27	989	30.00000	0.00000	29.800
"endz"	26	133	28.90000	0.00000	29.700
"killy"	27	1614	30.00000	0.10000	29.200
"mez"	27	1787	30.00000	0.10000	29.000
"bredrin"	25	79	27.80000	0.00000	28.700
"bossman"	26	1201	28.90000	0.00000	28.500
"playlist"	130	105877	144.30000	4.10000	28.500
"ahh"	49	24411	54.40000	0.90000	28.500
"deja"	35	10340	38.90000	0.40000	28.500
"shorty"	38	13501	42.20000	0.50000	28.400
"skank"	28	3472	31.10000	0.10000	28.300
"reload"	76	52815	84.40000	2.00000	28.000
"deh"	29	4996	32.20000	0.20000	27.800
"ragga"	25	991	27.80000	0.00000	27.700
"jaykae"	24	19	26.60000	0.00000	27.600
"sneakbo"	24	40	26.60000	0.00000	27.600
"gangsta"	34	10636	37.70000	0.40000	27.400
"wizkid"	25	1308	27.80000	0.10000	27.400
"nuff"	30	6880	33.30000	0.30000	27.100
"brum"	26	2743	28.90000	0.10000	27.000
"kno"	29	6066	32.20000	0.20000	26.900
"madness"	165	152856	183.20000	5.90000	26.600

Appendix 2. Keywords of tweets and replies -subcorpus (Reference corpus: EnTenTen)

Item	Frequency (focus)	Frequency (reference)	Relative frequency (focus)	Relative frequency (reference)	Score
"skepta"	252	1261	455.20000	0.00000	434.900
"loool"	220	250	397.40000	0.00000	394.600
"bro"	851	91441	1537.10000	3.50000	338.700
"grime"	374	29557	675.50000	1.10000	315.500
"jme"	183	1484	330.50000	0.10000	313.500
"fam"	366	32492	661.10000	1.30000	293.200
"giggs"	194	5411	350.40000	0.20000	290.500

"ting"	396	40896	715.30000	1.60000	277.200
"looool"	153	155	276.40000	0.00000	275.700
"bruv"	135	454	243.80000	0.00000	240.600
"rt"	1028	185342	1856.80000	7.20000	227.200
"loooool"	125	89	225.80000	0.00000	226.000
"kmt"	167	9989	301.60000	0.40000	218.200
"lol"	2217	451579	4004.50000	17.50000	216.700
"lool"	114	489	205.90000	0.00000	203.100
"dizzee"	114	1362	205.90000	0.10000	196.500
"looooool"	105	56	189.70000	0.00000	190.200
"coz"	167	20627	301.60000	0.80000	168.300
"nah"	236	40943	426.30000	1.60000	165.300
"riddim"	95	1761	171.60000	0.10000	161.600
"frisco"	147	16849	265.50000	0.70000	161.300
"brudda"	88	153	159.00000	0.00000	159.000
"ghetts"	86	112	155.30000	0.00000	155.700
"lotm"	86	940	155.30000	0.00000	150.800
"defo"	83	2371	149.90000	0.10000	138.200
"grm"	81	2634	146.30000	0.10000	133.700
"loooooool"	71	24	128.20000	0.00000	129.100
"ffs"	104	12949	187.90000	0.50000	125.800
"banger"	110	15578	198.70000	0.60000	124.600
"oi"	140	28558	252.90000	1.10000	120.600
"hahaha"	123	22470	222.20000	0.90000	119.300
"bizzle"	66	497	119.20000	0.00000	117.900
"jammer"	105	16075	189.70000	0.60000	117.500
"loll"	79	6742	142.70000	0.30000	113.900
"haha"	289	95689	522.00000	3.70000	111.200
"tbh"	102	17581	184.20000	0.70000	110.200
"luv"	99	18620	178.80000	0.70000	104.500
"thanku"	58	697	104.80000	0.00000	103.000
"looooooool"	55	17	99.30000	0.00000	100.300
"tbf"	61	2914	110.20000	0.10000	99.900
"fuckin"	110	26613	198.70000	1.00000	98.300
"insta"	66	5757	119.20000	0.20000	98.300
"stormzy"	57	2198	103.00000	0.10000	95.800
"ppl"	146	45559	263.70000	1.80000	95.800
"vibes"	126	37498	227.60000	1.50000	93.200
"tryna"	63	7514	113.80000	0.30000	88.900
"rascal"	90	22593	162.60000	0.90000	87.200
"mandem"	48	160	86.70000	0.00000	87.200
"ur"	250	109941	451.60000	4.30000	86.100
"nizzle"	45	113	81.30000	0.00000	81.900

"newham"	58	7924	104.80000	0.30000	80.900
"jheeze"	44	13	79.50000	0.00000	80.400
"hun"	101	33620	182.40000	1.30000	79.700
"dem"	181	82294	326.90000	3.20000	78.300
"naa"	64	13851	115.60000	0.50000	75.900
"looooooooool"	41	13	74.10000	0.00000	75.000
"hahahaha"	51	6842	92.10000	0.30000	73.600
"bday"	49	5794	88.50000	0.20000	73.100
"mixtape"	63	14813	113.80000	0.60000	73.000
"ima"	69	18704	124.60000	0.70000	72.900
"sn1"	41	1207	74.10000	0.00000	71.700
"smh"	60	13805	108.40000	0.50000	71.300
"kano"	65	18176	117.40000	0.70000	69.500
"omg"	141	70371	254.70000	2.70000	68.600
"yeh"	71	22948	128.20000	0.90000	68.400
"wtf"	101	43872	182.40000	1.70000	68.000
"bruva"	36	35	65.00000	0.00000	65.900
"remix"	119	59506	214.90000	2.30000	65.400
"facking"	36	271	65.00000	0.00000	65.300
"skint"	38	1808	68.60000	0.10000	65.100
"rah"	46	7549	83.10000	0.30000	65.100
"tho"	191	111768	345.00000	4.30000	64.900
"hahahahaha"	39	2617	70.40000	0.10000	64.900
"ya"	406	267227	733.30000	10.30000	64.700
"init"	75	28701	135.50000	1.10000	64.600
"wid"	50	10896	90.30000	0.40000	64.200
"congrats"	124	65520	224.00000	2.50000	63.600
"gunna"	40	4022	72.30000	0.20000	63.400
"nang"	51	12129	92.10000	0.50000	63.400
"hahah"	38	3042	68.60000	0.10000	62.300
"thankyou"	70	27383	126.40000	1.10000	61.900
"gonna"	599	430275	1081.90000	16.70000	61.300
"wanna"	278	189372	502.10000	7.30000	60.400
"mwah"	34	926	61.40000	0.00000	60.300
"xmas"	69	29506	124.60000	1.10000	58.600
"lolll"	32	112	57.80000	0.00000	58.500
"hahahah"	33	1291	59.60000	0.00000	57.700
"dat"	108	62195	195.10000	2.40000	57.500
"innit"	34	2579	61.40000	0.10000	56.700
"xx"	143	95243	258.30000	3.70000	55.300
"itscleopatra"	30	0	54.20000	0.00000	55.200
"lmao"	50	16922	90.30000	0.70000	55.200
"yea"	148	100295	267.30000	3.90000	54.900

"freestyle"	100	60514	180.60000	2.30000	54.300
"lmfao"	33	3283	59.60000	0.10000	53.800
"pls"	47	15753	84.90000	0.60000	53.400
"truss"	85	51070	153.50000	2.00000	51.900
"bluku"	28	5	50.60000	0.00000	51.600
"rudeboy"	28	257	50.60000	0.00000	51.100
"nigga"	83	51252	149.90000	2.00000	50.600
"babe"	180	141205	325.10000	5.50000	50.400
"fav"	52	22821	93.90000	0.90000	50.400
"dunno"	76	45062	137.30000	1.70000	50.400
"jollof"	28	625	50.60000	0.00000	50.400
"looooooooool"	27	10	48.80000	0.00000	49.700
"u"	1669	1570125	3014.60000	60.80000	48.800
"mate"	477	433301	861.60000	16.80000	48.500
"duppy"	27	674	48.80000	0.00000	48.500
"dawg"	37	10321	66.80000	0.40000	48.500
"gotta"	210	179204	379.30000	6.90000	47.900
"badman"	28	1980	50.60000	0.10000	47.900
"mad"	457	420404	825.50000	16.30000	47.800
"deffo"	27	1102	48.80000	0.00000	47.700
"vibez"	26	154	47.00000	0.00000	47.700
"pre-order"	75	48826	135.50000	1.90000	47.200
"feat"	201	173487	363.10000	6.70000	47.200
"cmon"	29	3416	52.40000	0.10000	47.100
"mc"	186	159389	336.00000	6.20000	47.000
"yizzy"	25	6	45.20000	0.00000	46.100
"icymi"	26	1410	47.00000	0.10000	45.500
"shit"	441	433754	796.60000	16.80000	44.800
"ragga"	25	991	45.20000	0.00000	44.500
"ft"	355	354039	641.20000	13.70000	43.700
"adz"	24	740	43.40000	0.00000	43.100
"hoodies"	24	1083	43.40000	0.00000	42.600
"ep"	192	185273	346.80000	7.20000	42.500
"jheeeze"	23	1	41.50000	0.00000	42.500
"bredrin"	23	79	41.50000	0.00000	42.400
"tonight"	535	565574	966.30000	21.90000	42.200
"merch"	31	9069	56.00000	0.40000	42.200
"holla"	26	3804	47.00000	0.10000	41.800
"deh"	27	4996	48.80000	0.20000	41.700
"gyal"	24	1748	43.40000	0.10000	41.500
"hus"	32	11014	57.80000	0.40000	41.200
"spotify"	79	64517	142.70000	2.50000	41.100
"cleosnaps"	22	0	39.70000	0.00000	40.700

"headie"	22	77	39.70000	0.00000	40.600
"ahh"	43	24411	77.70000	0.90000	40.400
"cz"	37	17952	66.80000	0.70000	40.000
"mazza"	24	2971	43.40000	0.10000	39.800
"gona"	24	3012	43.40000	0.10000	39.700
"footsie"	23	2099	41.50000	0.10000	39.300
"wiley"	96	89413	173.40000	3.50000	39.100
"woooah"	21	68	37.90000	0.00000	38.800
"aint"	51	36349	92.10000	1.40000	38.700
"lockdown"	35	17299	63.20000	0.70000	38.500
"aj"	81	73260	146.30000	2.80000	38.400
"wamp"	22	1643	39.70000	0.10000	38.300
"comming"	27	7978	48.80000	0.30000	38.000
"woi"	21	654	37.90000	0.00000	38.000
"bbk"	22	1982	39.70000	0.10000	37.800
"woah"	28	9407	50.60000	0.40000	37.800
"aww"	31	13251	56.00000	0.50000	37.700
"yo"	100	99066	180.60000	3.80000	37.600
"yute"	21	989	37.90000	0.00000	37.500
"buss"	43	28593	77.70000	1.10000	37.300
"aswell"	36	19871	65.00000	0.80000	37.300
"yessss"	21	1213	37.90000	0.00000	37.200
"bludklart"	20	0	36.10000	0.00000	37.100
"omds"	20	63	36.10000	0.00000	37.000
"dere"	24	5586	43.40000	0.20000	36.500
"geezer"	26	8463	47.00000	0.30000	36.100
"skank"	22	3472	39.70000	0.10000	35.900
"gigg"	20	1065	36.10000	0.00000	35.700
"wizkid"	20	1308	36.10000	0.10000	35.300
"gwan"	19	932	34.30000	0.00000	34.100
"skep"	19	1014	34.30000	0.00000	34.000
"gangsta"	26	10636	47.00000	0.40000	34.000
"nuh"	24	7921	43.40000	0.30000	33.900
"pl"	120	141146	216.80000	5.50000	33.700
"convo"	23	6809	41.50000	0.30000	33.700
"yana"	21	4457	37.90000	0.20000	33.200
"vid"	46	39671	83.10000	1.50000	33.200
"kno"	22	6066	39.70000	0.20000	33.000
"aswel"	18	447	32.50000	0.00000	32.900
"prod"	53	50418	95.70000	2.00000	32.800
"retweet"	28	14906	50.60000	0.60000	32.700
"chale"	18	791	32.50000	0.00000	32.500
"tweet"	310	419954	559.90000	16.30000	32.500

"sis"	86	99255	155.30000	3.80000	32.300
"brodie"	30	18540	54.20000	0.70000	32.100
"vex"	33	23167	59.60000	0.90000	31.900
"jackuum"	17	3	30.70000	0.00000	31.700
"kyze"	17	7	30.70000	0.00000	31.700
"raskit"	17	19	30.70000	0.00000	31.700
"sneakbo"	17	40	30.70000	0.00000	31.700
"gooden"	19	3052	34.30000	0.10000	31.600
"yooo"	17	165	30.70000	0.00000	31.500
"fredo"	18	1745	32.50000	0.10000	31.400
"yout"	18	1901	32.50000	0.10000	31.200
"nuff"	21	6880	37.90000	0.30000	30.700
"throwback"	34	26731	61.40000	1.00000	30.700
"ic3"	18	2425	32.50000	0.10000	30.600
"diss"	27	16283	48.80000	0.60000	30.500
"fuck"	575	856316	1038.60000	33.20000	30.400
"clout"	40	36578	72.30000	1.40000	30.300
"shorty"	25	13501	45.20000	0.50000	30.300
"pree"	17	1211	30.70000	0.00000	30.300
"alie"	17	1228	30.70000	0.00000	30.300
"fekky"	16	33	28.90000	0.00000	29.900

Appendix 3. Keywords of the retweets subcorpus (Reference corpus: EnTenTen)

Item	Frequency (focus)	Frequency (reference)	Relative frequency (focus)	Relative frequency (reference)	Score
"rt"	16480	185342	47486.40000	7.20000	5807.600
"ghetts"	252	112	726.10000	0.00000	724.000
"grime"	508	29557	1463.80000	1.10000	683.000
"skepta"	197	1261	567.60000	0.00000	542.200
"kano"	308	18176	887.50000	0.70000	521.500
"frisco"	246	16849	708.80000	0.70000	429.600
"hoodies"	149	1083	429.30000	0.00000	413.000
"banger"	223	15578	642.60000	0.60000	401.400
"ghett"	121	35	348.70000	0.00000	349.200
"dizzee"	113	1362	325.60000	0.10000	310.200
"grm"	114	2634	328.50000	0.10000	299.000
"icymi"	98	1410	282.40000	0.10000	268.700
"cleo"	135	13252	389.00000	0.50000	257.700
"bratt"	80	3227	230.50000	0.10000	205.800
"newham"	89	7924	256.40000	0.30000	197.000
"jackuum"	62	3	178.70000	0.00000	179.600
"giggs"	74	5411	213.20000	0.20000	177.100

"nutcrackerz"	58	2	167.10000	0.00000	168.100
"riddim"	60	1761	172.90000	0.10000	162.800
"mz"	77	10756	221.90000	0.40000	157.300
"rascal"	99	22593	285.30000	0.90000	152.700
"insta"	60	5757	172.90000	0.20000	142.200
"ting"	121	40896	348.70000	1.60000	135.300
"jools"	51	3206	147.00000	0.10000	131.600
"prez"	55	6926	158.50000	0.30000	125.800
"bluku"	43	5	123.90000	0.00000	124.900
"mixtape"	67	14813	193.10000	0.60000	123.300
"vibes"	93	37498	268.00000	1.50000	109.700
"jme"	38	1484	109.50000	0.10000	104.500
"sparko"	34	93	98.00000	0.00000	98.600
"prod"	97	50418	279.50000	2.00000	95.000
"popcaan"	33	506	95.10000	0.00000	94.200
"remix"	104	59506	299.70000	2.30000	91.000
"bizzle"	31	497	89.30000	0.00000	88.600
"bro"	139	91441	400.50000	3.50000	88.400
"ep"	249	185273	717.50000	7.20000	87.900
"tix"	36	5013	103.70000	0.20000	87.700
"ten10"	29	101	83.60000	0.00000	84.200
"devlin"	46	15821	132.50000	0.60000	82.800
"raskit"	28	19	80.70000	0.00000	81.600
"bbk"	30	1982	86.40000	0.10000	81.200
"adz"	28	740	80.70000	0.00000	79.400
"pan-fried"	31	4142	89.30000	0.20000	77.800
"oi"	54	28558	155.60000	1.10000	74.400
"feat"	197	173487	567.60000	6.70000	73.700
"freestyle"	83	60514	239.20000	2.30000	71.800
"tonight"	557	565574	1605.00000	21.90000	70.100
"mc"	171	159389	492.70000	6.20000	68.800
"lotm"	24	940	69.20000	0.00000	67.700
"teardrop"	34	12596	98.00000	0.50000	66.500
"collab"	30	8725	86.40000	0.30000	65.400
"album"	1327	1542218	3823.70000	59.70000	63.000
"duppy"	22	674	63.40000	0.00000	62.800
"ft"	310	354039	893.30000	13.70000	60.800
"fris"	21	410	60.50000	0.00000	60.500
"nuh"	27	7921	77.80000	0.30000	60.300
"fav"	38	22821	109.50000	0.90000	58.700
"vid"	51	39671	147.00000	1.50000	58.300
"sbtv"	20	251	57.60000	0.00000	58.100
"retweet"	31	14906	89.30000	0.60000	57.300

"footsie"	21	2099	60.50000	0.10000	56.900
"kojo"	21	2166	60.50000	0.10000	56.800
"defo"	21	2371	60.50000	0.10000	56.300
"nah"	50	40943	144.10000	1.60000	56.100
"brudda"	19	153	54.70000	0.00000	55.400
"reload"	56	52815	161.40000	2.00000	53.300
"dubplate"	18	370	51.90000	0.00000	52.100
"ic3"	19	2425	54.70000	0.10000	51.000
"deja"	24	10340	69.20000	0.40000	50.100
"dem"	72	82294	207.50000	3.20000	49.800
"bonkers"	26	14064	74.90000	0.50000	49.200
"rapper"	77	91670	221.90000	3.50000	49.000
"fam"	38	32492	109.50000	1.30000	48.900
"lockdown"	28	17299	80.70000	0.70000	48.900
"smash"	157	217586	452.40000	8.40000	48.100
"revvin"	16	113	46.10000	0.00000	46.900
"jammer"	26	16075	74.90000	0.60000	46.800
"ldn"	20	6682	57.60000	0.30000	46.600
"playlist"	82	105877	236.30000	4.10000	46.500
"pre-order"	45	48826	129.70000	1.90000	45.200
"manor"	111	160158	319.80000	6.20000	44.600
"waze"	18	4978	51.90000	0.20000	44.300
"f64"	15	268	43.20000	0.00000	43.800
"bruv"	15	454	43.20000	0.00000	43.500
"gigg"	15	1065	43.20000	0.00000	42.500
"unreal"	45	53696	129.70000	2.10000	42.400
"i_d"	14	86	40.30000	0.00000	41.200
"congrats"	50	65520	144.10000	2.50000	41.000
"bday"	17	5794	49.00000	0.20000	40.800
"stacks"	18	7642	51.90000	0.30000	40.800
"stormzy"	15	2198	43.20000	0.10000	40.800
"hoodie"	30	31175	86.40000	1.20000	39.600
"bossman"	14	1201	40.30000	0.00000	39.500
"madness"	94	152856	270.90000	5.90000	39.300
"sym"	16	5594	46.10000	0.20000	38.700
"fave"	29	31192	83.60000	1.20000	38.300
"endz"	13	133	37.50000	0.00000	38.300
"rudeboy"	13	257	37.50000	0.00000	38.100
"yo"	63	99066	181.50000	3.80000	37.700
"whippin"	13	694	37.50000	0.00000	37.500
"brum"	14	2743	40.30000	0.10000	37.400
"flatmate"	15	5471	43.20000	0.20000	36.500
"killy"	13	1614	37.50000	0.10000	36.200

"mcs"	26	28946	74.90000	1.10000	35.800
"badman"	13	1980	37.50000	0.10000	35.700
"spotify"	43	64517	123.90000	2.50000	35.700
"lyrical"	42	62988	121.00000	2.40000	35.500
"inferno"	27	31597	77.80000	1.20000	35.400
"mandem"	12	160	34.60000	0.00000	35.400
"jubbly"	12	203	34.60000	0.00000	35.300
"brixx"	12	210	34.60000	0.00000	35.300
"mics"	15	6600	43.20000	0.30000	35.200
"merch"	16	9069	46.10000	0.40000	34.900
"da"	172	342406	495.60000	13.30000	34.800
"mad"	206	420404	593.60000	16.30000	34.400
"tropez"	14	5451	40.30000	0.20000	34.100
"deffo"	12	1102	34.60000	0.00000	34.100
"crb"	17	12343	49.00000	0.50000	33.800
"wamp"	12	1643	34.60000	0.10000	33.400
"rap"	89	173902	256.40000	6.70000	33.300
"chippy"	13	4364	37.50000	0.20000	32.900
"aint"	27	36349	77.80000	1.40000	32.700
"jaykae"	11	19	31.70000	0.00000	32.700
"mazza"	12	2971	34.60000	0.10000	31.900
"wavey"	11	675	31.70000	0.00000	31.900
"maida"	13	5959	37.50000	0.20000	31.200
"novelist"	54	105226	155.60000	4.10000	30.900
"mez"	11	1787	31.70000	0.10000	30.600
"woah"	14	9407	40.30000	0.40000	30.300
"brixton"	16	14340	46.10000	0.60000	30.300
"gig"	89	195599	256.40000	7.60000	30.000
"mosh"	12	4929	34.60000	0.20000	29.900
"kmt"	14	9989	40.30000	0.40000	29.800
"skengman"	10	7	28.80000	0.00000	29.800
"ash"	172	406206	495.60000	15.70000	29.700
"thug"	42	80568	121.00000	3.10000	29.600
"ig"	34	60696	98.00000	2.40000	29.500
"loool"	10	250	28.80000	0.00000	29.500
"vibe"	56	116681	161.40000	4.50000	29.400
"glasto"	10	528	28.80000	0.00000	29.200
"premiere"	92	212412	265.10000	8.20000	28.800
"tune"	333	834208	959.50000	32.30000	28.800
"insomnia"	47	97730	135.40000	3.80000	28.500
"gyal"	10	1748	28.80000	0.10000	27.900
"shoutout"	11	4539	31.70000	0.20000	27.800
"ashes"	12	7844	34.60000	0.30000	27.300

"tracklist"	11	5186	31.70000	0.20000	27.200
"ur"	49	109941	141.20000	4.30000	27.000
"wid"	13	10896	37.50000	0.40000	27.000
"dere"	11	5586	31.70000	0.20000	26.900
"og"	22	36689	63.40000	1.40000	26.600
"lool"	9	489	25.90000	0.00000	26.400
"sidewinder"	11	6403	31.70000	0.20000	26.200
"realness"	10	3811	28.80000	0.10000	26.000
"acen"	9	973	25.90000	0.00000	26.000
"castella"	9	1113	25.90000	0.00000	25.800
"yessss"	9	1213	25.90000	0.00000	25.700
"wanna"	73	189372	210.30000	7.30000	25.400
"shorty"	13	13501	37.50000	0.50000	25.300
"ticket"	495	1441461	1426.30000	55.80000	25.100
"ibiza"	15	19969	43.20000	0.80000	24.900
"gonna"	151	430275	435.10000	16.70000	24.700
"sittin"	10	5424	28.80000	0.20000	24.600
"preorder"	12	11860	34.60000	0.50000	24.400
"ghetto"	31	70023	89.30000	2.70000	24.300
"lil"	32	73311	92.20000	2.80000	24.300
"congratulations"	16	24512	46.10000	0.90000	24.200
"hmv"	10	6085	28.80000	0.20000	24.100
"eskimo"	16	24737	46.10000	1.00000	24.100
"i_skream"	8	0	23.10000	0.00000	24.100
"sanelly"	8	1	23.10000	0.00000	24.100
"ep30"	8	10	23.10000	0.00000	24.000
"jheeze"	8	13	23.10000	0.00000	24.000
"askem"	8	38	23.10000	0.00000	24.000
"headie"	8	77	23.10000	0.00000	24.000
"hyp"	13	15734	37.50000	0.60000	1,583333333
"ss19"	8	343	23.10000	0.00000	1,444444444
"grimey"	8	600	23.10000	0.00000	1,305555556
"digga"	8	626	23.10000	0.00000	1,305555556
"yardie"	8	629	23.10000	0.00000	1,305555556
"moonchild"	8	688	23.10000	0.00000	1,236111111
"aswell"	14	19871	40.30000	0.80000	1,236111111
"salute"	42	109371	121.00000	4.20000	1,166666667
"parklife"	8	900	23.10000	0.00000	1,097222222
"capo"	12	14057	34.60000	0.50000	23.000
"goosebumps"	9	4397	25.90000	0.20000	23.000
"konan"	8	1369	23.10000	0.10000	1,472222222
"nite"	14	21151	40.30000	0.80000	1,402777778
"omg"	29	70371	83.60000	2.70000	1,402777778

"bloggs"	8	1579	23.10000	0.10000	1,402777778
"nme"	10	8349	28.80000	0.30000	1,263888889
"dat"	26	62195	74.90000	2.40000	1,125
"pls"	12	15753	34.60000	0.60000	0,986111111
"legend"	182	596146	524.40000	23.10000	1,430555556
"dope"	38	105147	109.50000	4.10000	1,430555556
"imma"	9	6316	25.90000	0.20000	1,291666667
"lewisham"	10	10152	28.80000	0.40000	1,152777778
"nuff"	9	6880	25.90000	0.30000	1,083333333
"skengdo"	7	3	20.20000	0.00000	1,013888889
"boobeam"	7	4	20.20000	0.00000	1,013888889