

# Race, racism and race equality in the UK press

*Carmen Dayrell and Elena Semino*

## 1. Introduction

This study investigates how the UK press frames race issues, and racism and race equality in particular. The analysis focuses on articles published between 01.10.2018 and 31.09.2019 by all major British broadsheet and tabloid newspapers (see details in section 2). Our goal is to uncover the amount of attention that race issues have received in the press and the most dominant narratives around them.

## 2. Data

The analysis is drawn from three separate datasets, or ‘corpora’, each consisting of two subcorpora: a set of broadsheet articles and a set of tabloid articles. The focus is on articles published between 01.10.2018 and 31.09.2019 by the following newspaper titles, in their weekday and Sunday versions:

Broadsheet newspapers:

- The Guardian / The Observer
- The Times / The Sunday Times
- The Daily Telegraph / The Sunday Telegraph
- The Independent
- i

Tabloids:

- The Daily Mail / Mail on Sunday
- The Daily Mirror
- The Sun
- The Express / Sunday Express
- The Daily Star / Daily Star Sunday
- The Metro

The collection of individual texts proceeded on the basis of specific terms, as detailed below. Each corpus was named after the most relevant term.

- The **Grime corpus** is made up of texts containing the word ‘grime’;
- The **Stop-and-Search corpus** comprises texts containing either ‘stop and search(es)’ or the hyphenised phrase ‘stop-and-search(es)’;
- The **Race Equality corpus** includes texts containing one of the following: ‘race (in)equality’, ‘racial (in)equality’, ‘racial (dis)parity’, or ‘racial (in)justice’.

In the case of the Race Equality corpus, we started by searching for texts that contained the phrase ‘race equality’ specifically. However, as the resulting number of texts was very small, particularly in the tabloids (102 texts from the broadsheet collection and 7 texts from the tabloids), we extended the search to

other related terms as an attempt to gather a higher number of texts. The resulting number of texts is presented below. Section 2 will show the frequency of each term in each subcorpus.

All texts were collected from a news aggregator service (*LexisNexis*), considering the printed version of the newspapers. We selected the national editions only, thus excluding the Irish, Scottish and Northern Ireland editions as well as newswires and websites. As for duplicates, we removed identical texts, but kept those with moderate similarity, such as similar articles published on different days or by different newspapers. In terms of genre, in addition to news reports, we also included editorials and letters to the editor.

Table 1 shows the number of texts and total number of words in each corpus. This serves as an indication of the amount of attention that the UK press paid to race issues during the period under analysis. Note that the number of texts is consistently higher in the broadsheet newspapers as compared to the number of texts published by the tabloids. In other words, broadsheet newspapers discussed race issues using our search terms more often than the tabloids; this difference was especially evident in the Race Equality corpus, where the number of texts in the broadsheets outnumbered those in the tabloids by more than 10 times.

Corpus	Broadsheet		Tabloids	
	Number of texts	Number of Words	Number of texts	Number of Words
The Grime corpus	648	742,477	304	134,787
The Stop-and-Search corpus	526	595,287	341	181,789
The Race Equality corpus	468	592,420	41	29,097

Table 1: Number of texts and total number of words comprising each corpus

In what follows, we present the distribution of articles across the various newspaper titles within each corpus, considering the broadsheet and the tabloids subsets separately. Figures 1 and 2 show the figures for the Grime corpus, Figures 3 and 4 indicate those in the Stop-and-Search corpus, and Figures 5 and 6 show the number of texts across the newspaper titles in the Race Equality corpus.

**Grime Corpus: Number of articles in broadsheet newspapers**

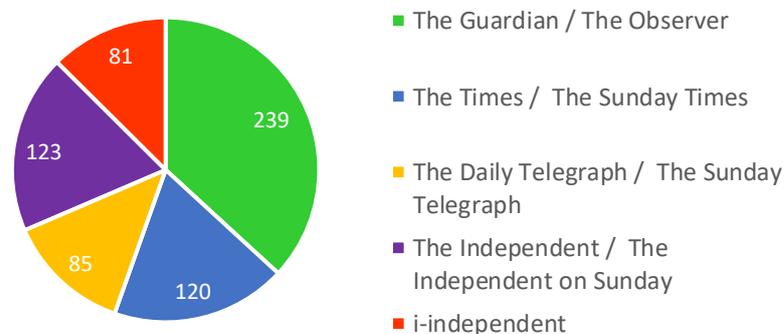


Figure 1: Number of texts from each broadsheet titles in the Grime Corpus

**Grime Corpus: Number of texts in the tabloids**

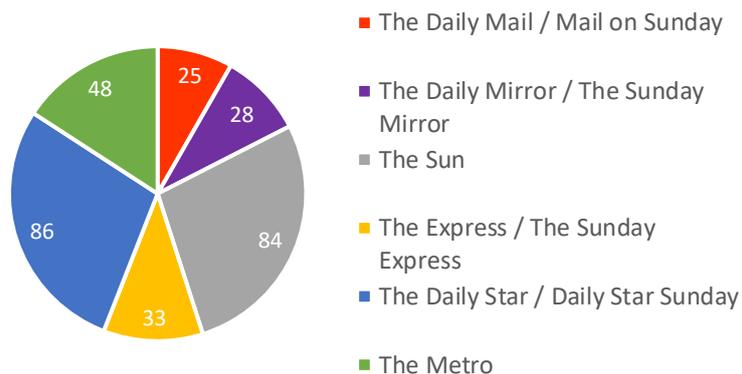


Figure 2: Number of texts from each tabloid titles in the Grime Corpus

**Stop-and-Search Corpus:  
Number of texts in broadsheet newspapers**

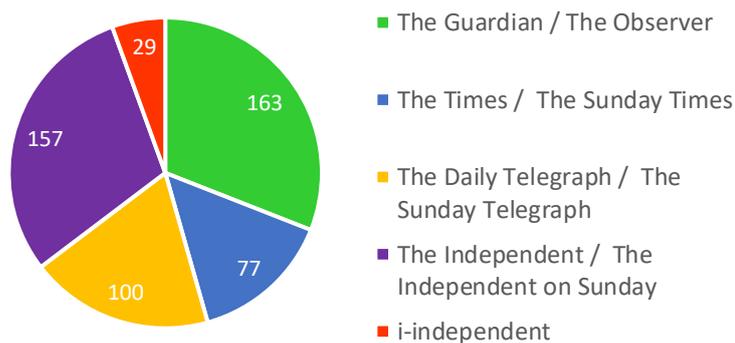


Figure 3: Number of texts from each broadsheet titles in the Stop-and-Search Corpus

**Stop-and-Search Corpus:  
Number of texts in the tabloids**

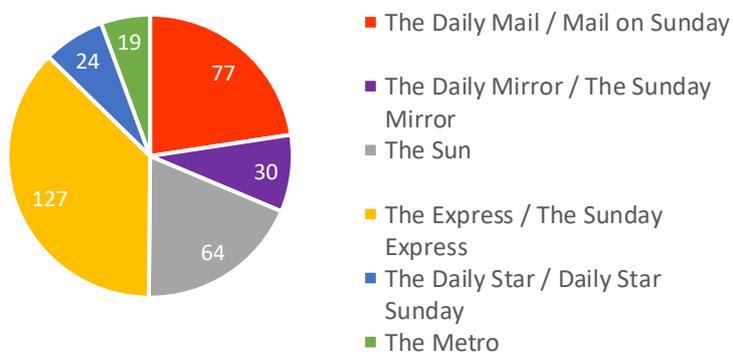


Figure 4: Number of texts from each tabloid titles in the Stop-and-Search Corpus

**Race Equality Corpus:  
Number of texts in broadsheet newspapers**

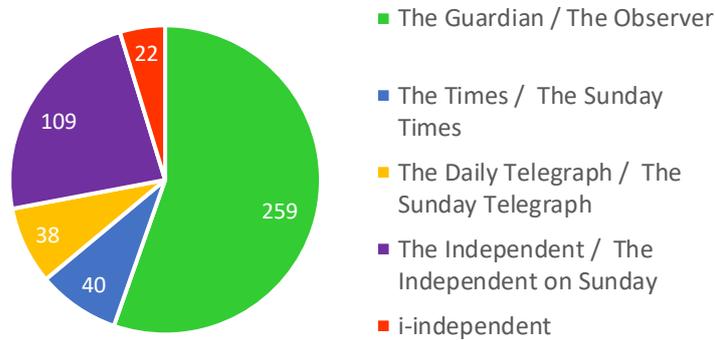


Figure 5: Number of texts from each broadsheet titles in the Race Equality Corpus

**Race Equality Corpus:  
Number of texts in the tabloids**

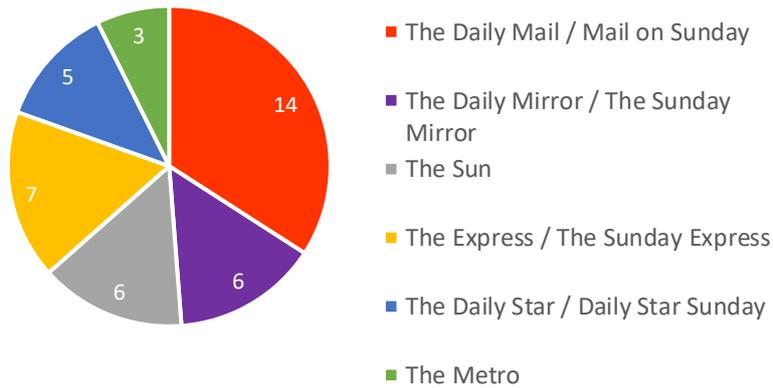


Figure 6: Number of texts from each tabloid title in the Race Equality Corpus

As can be seen, *The Guardian* accounts for the highest number of articles in all three broadsheet subcorpora: 37% of the Grime broadsheet subcorpus (Figure 1), 31% of the Stop-and-Search broadsheet (Figure 3) and as much as 55% of the Race Equality broadsheet (Figure 5). *The Independent* comes second, with about 20% in the Grime and Race Equality broadsheet subcorpora and, like *The Guardian*, about a third of the texts in the Stop-and-Search broadsheet.

Within the tabloids, different newspaper titles seem to focus on different topics. In the Grime tabloid subcorpus, *The Sun* and *The Daily Star* account for the largest number of articles, with a third of the articles each (Figure 2). In the Stop-and-Search tabloid corpus, as much as 37% of texts come from *The Express*, and *The Daily Mail* comes second with 22% (Figure 4). As for the Race Equality tabloids corpus (Figure 6), *The Daily Mail* accounts for the largest number of texts (34%).

### 3. Distribution of words across time

We now turn to look at the distribution of articles about each of our topics across time by plotting the total number of words per month in each corpus for the broadsheet and tabloids subcorpora separately, considering all the newspaper titles combined (Figures 7 and 8 respectively). These figures give us an indication of how the level of attention fluctuated in the UK press throughout time

. Note that the scale varies from one corpus to another.

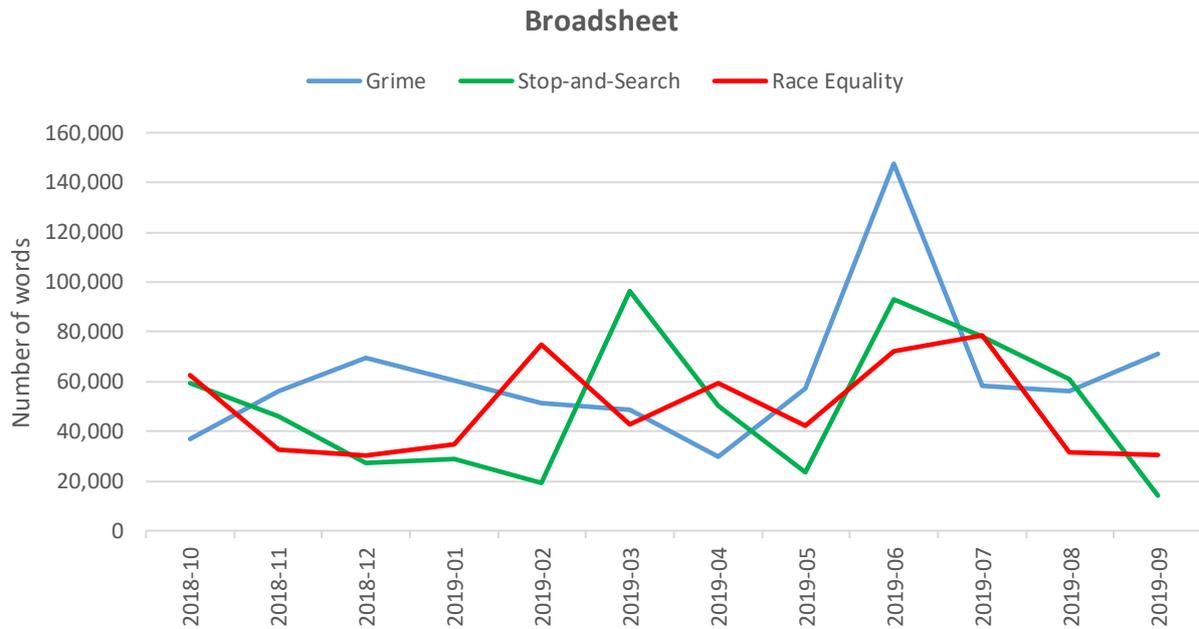


Figure 7: Distribution of words in broadsheet newspapers across time

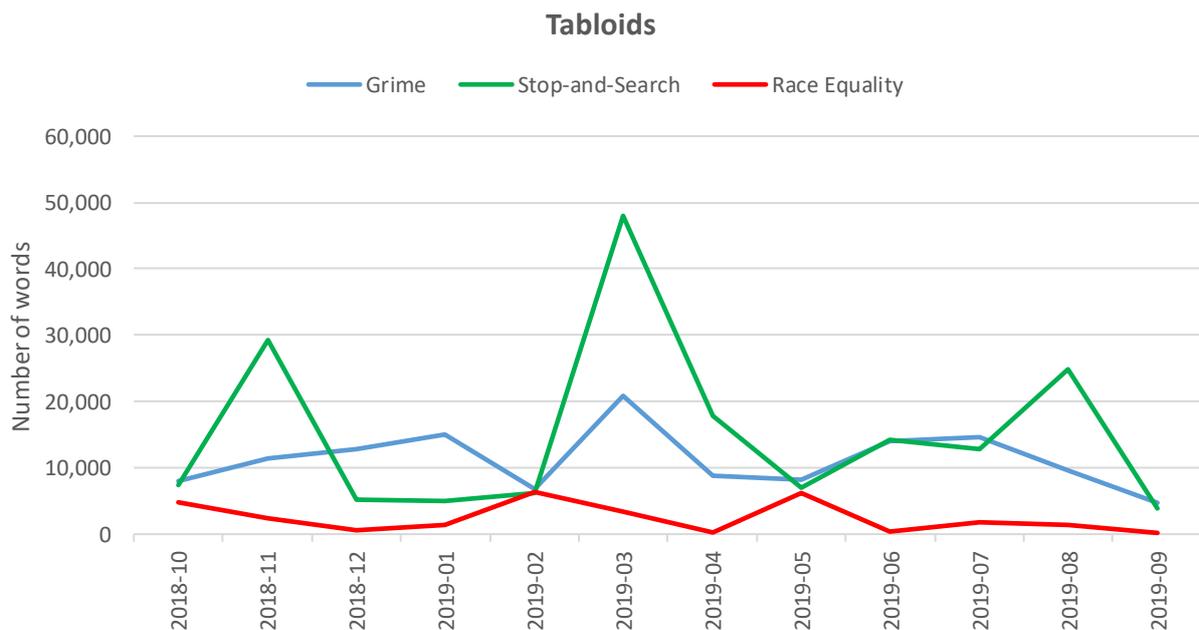


Figure 8: Distribution of words in the tabloids across time

As can be seen, the number of words peaked in different months in the broadsheet and tabloids, irrespective of the corpus. In other words, this seems to suggest that, with very few exceptions, racial issues did not attract the attention of broadsheet and tabloids around the same time. The exceptions were the peaks in the broadsheet and the tabloids Stop-and-Search subcorpora in Mar/2019 and in the Race Equality corpus in Feb/2019. A preliminary analysis of the articles – by reading the concordance lines of the query terms during each peak – pointed towards some salient themes discussed in the newspapers at those specific points in time. These are discussed briefly in this section. Further evidence is provided when we examine each corpus in more detail (sections 6 to 8).

The Grime broadsheet subcorpus had its highest peak in Jun/2019 whereas the tabloids showed a less pronounced peak and it lasted for a slightly longer period (Jun and Jul/2019). Both peaks were triggered by the reporting on the performance of the British rapper Stormzy on the Pyramid stage at the Glastonbury festival on 28 June 2019. The broadsheet newspapers also showed a minor peak in Dec/2018 due to an increase in articles about albums of the year. In the tabloids, a minor peak was seen a month after, in Jan/2019. This was mainly due to reports on two incidents at the time: (i) a 21-year-old male who drove his 4x4 car onto the dancefloor of a nightclub in London, causing serious injuries to a number of people who were waiting to see the grime rap artist Giggs and (ii) the imprisonment of grime star Ollie Millington for the rape of four women. The Grime tabloid subcorpus also peaked slightly in Mar/2019 as a result of reports on the death of Louella Fletcher-Michie at Bestival music festival in Dorset due to an overdose of drugs supplied by her boyfriend.

In the Stop-and-Search corpora, as mentioned earlier, both the broadsheet and the tabloids peaked in Mar/2019. Although there were peaks in other months, these peaks did not occur simultaneously in the broadsheets and in the tabloids. The broadsheet subcorpus peaked in Jun/2019 and the number of words remained high until Aug/2019, with the figure dropping sharply in Sep/2019. Similar to the pattern shown in the broadsheet, the number of words in the tabloids also started to rise in Jun/2019; however, the peak only occurred in Aug/2019. The tabloid subcorpus also peaked in Nov/2018. All these peaks related to reports about funding for the police force in both the broadsheets and the tabloids. While some articles reported on the extra funding allocated to police services, others mentioned a shortage of police officers, as a result of underfunding. The broadsheets also highlighted an increase in crime rates in England, and stabbings in particular, reporting on a heated discussion around the Conservative party's proposal to widen police stop-and-search powers to help prevent those crimes. The tabloids reported on young males being victims of crimes across England, many stabbed to death.

As for the Race Equality corpus, in addition to the peak in Feb/2019 in both the broadsheets and the tabloids, the broadsheet also showed a minor peak in Apr/2019 while the tabloids peaked in the following month (May/2019). The broadsheet peaked again even higher in Jun/2019, a peak that prevailed in Jul/2019, and the number of words fell to the lowest levels after that. Peaks in the broadsheets were due to articles discussing institutional racism and racial discrimination in the UK as well as in the US. For the latter, the discussion revolved around accusations against Trump and discrimination in the American Football League. In the tabloids, the peak in Feb/2019 was related to articles containing film synopses which mentioned the director Barry Jenkins and *Moonlight*, the winner of Best Picture Oscar in 2017 that addressed racial injustice. The peak in the tabloids in May/2019 was due to a long article by *The Daily Mail* which reported on racial tensions in Oldham between the Asian and the white British communities.

#### 4. Frequency of individual terms in the corpora

Table 2 shows the raw and relative frequencies (per 10,000 words) of the search terms used to collect the texts that comprise each corpus, considering the broadsheet and tabloid subcorpora separately. For the Stop-and-Search corpus, the table shows the combined frequencies of the terms ‘stop and search(es)’ and ‘stop-and-search(es)’.

Corpus	Search term	Broadsheet		Tabloids	
		Raw Frequency	Relative Frequency (per 10,000 words)	Raw Frequency	Relative Frequency (per 10,000 words)
Grime Corpus	grime	895	12.1	357	26.5
Stop-and-Search Corpus	stop and search(es)	1,266	21.3	596	32.8
Race Equality Corpus	race equality	102	1.7	7	2.4
	racial equality	130	2.2	21	7.2
	racial parity	3	0.1	0	0.0
	racial justice	117	2.0	2	0.7
	race inequality	5	0.1	0	0.0
	racial inequality	100	1.7	8	2.7
	racial disparity	33	0.6	0	0.0
	racial injustice	100	1.7	10	3.4

Table 2: Raw and relative frequencies of the search terms in each corpus

Note that the raw frequencies are consistently lower in the tabloid corpora. This is expected given that the tabloids contain a lower number of texts. Also note that most terms are proportionally more frequent in the tabloids than in the broadsheet newspapers. The exceptions are ‘racial justice’ and those terms with no occurrence in the tabloids (‘racial parity’ and ‘disparity’ and ‘race inequality’). This is due to the nature of the calculation of relative frequencies. As tabloid corpora contain fewer words, the resulting relative frequency is inevitably higher.

Thus, the table reflects the contribution of each search term for the composition of each subcorpus. This is especially relevant in the analysis of the Race Equality corpus. As can be seen, the press showed a clear preference for some terms over others.

#### 5. Methodology

We now explain how we used analytical techniques associated with the field of Corpus Linguistics to study the dominant narratives in each corpus, split by broadsheets and tabloids, and within each considering all newspaper titles in aggregate. To provide an overview of the most distinctive linguistic characteristics of each corpus, we carried out ‘keyword’ analyses. Keywords are words that are more frequent in a corpus of interest (known as the ‘study’ corpus) than they are in another corpus (known as the ‘reference corpus’), where the difference is statistically significant (McEnery and Hardie, 2011: 245). They can be interpreted as reflecting the most distinctive concepts and themes in a particular corpus. Keywords were calculated by means of the software tool *WMatrix* (Rayson 2008), using the 2006 British English corpus (BE2006) as the reference corpus – a general corpus of British English.

For the calculation of keywords, we established a minimum raw frequency of 20 occurrences in the BE2006. This was because there is a time difference between texts in the study corpora (2018-2019) and those in the BE2006 (2006). This means that words in the study corpus which occurred with a low frequency

in the BE2006 (such as politicians' proper names) would inevitably emerge as keywords. Relevant words discarded at this stage emerged in the qualitative analysis. For the study corpora, we considered those words with a minimum frequency of 30 occurrences per 100,000 words. It therefore means that the minimum raw frequency used for the calculation varied from one subcorpus to another. In terms of statistical tests, we applied the log-likelihood test, considering a critical value higher than 6.63 ( $p < 0.01$ ). The log-likelihood test tells us to what extent differences in frequencies between the two corpora (the study and the reference corpus) is statistically significant. Here, it was combined with log-ratio (Hardie *forthcoming*), an effect size measure which measures how big the difference is. The higher the LogRatio score, the larger the difference. The log-ratio calculation therefore gives us the words whose frequencies are proportionally higher in the study corpus. We considered a minimum log-ratio score of 3, that is, words that appeared at least eight times more frequently in the study corpus than in the BE2006.

As this research project seeks to gain insights into how the UK press discusses racial issues, especially with regard to BAME people, we then focused on words which could function as references to race. For the purposes of this study, we considered the following: 'ethnic minority(ies)', 'BAME', 'BME' as well as terms related to skin colour or nationality ('black', 'of colour', 'Asian', 'Indian', 'Pakistani', and 'Bangladeshi'). We also included occurrences of the word 'white', as a benchmark for comparison. In our analyses, we started by examining the frequencies of occurrence of each term in each subcorpus. For the two words that were found to be used most frequently – 'black' and 'white' – we carried out an additional 'collocation' analysis.

Collocation analyses explore co-occurrence relationships between words, and therefore makes it possible to study the narratives or discourses that a word is part of. A word collocates with another if it is 'more likely to occur in the presence of the other than elsewhere' (McEnery and Hardie, 2011: 240). Collocations were generated by means of the software package *LancsBox* (Brezina 2015 et al), on the basis of the following criteria:

- Span of 5:5 – a window of five words to the left and five words to the right of the search word;
- Mutual Information (MI) score  $\geq 6$ . MI is a statistical procedure widely employed in corpus studies to indicate how strong the association between two words is. It is calculated by considering their frequency of co-occurrence in relation to their frequencies when occurring independently in each corpus.
- Minimum frequency of collocation: a minimum of 10 instances in the broadsheet subcorpora and five instances in the tabloids.

The final step was to explore the collocations of these selected race-related words by close reading of their 'concordance lines', that is, individual occurrences of each word with the preceding and following stretches of text (McEnery and Hardie 2011: 241).

## **6. The Grime Corpus**

In this section, we discuss the results from the analysis of the Grime Corpus. We first present the keywords that met the established criteria (see section 5) in the Broadsheet and Tabloid Grime subcorpora and discuss the themes that emerged in each subcorpus. The second part of the section presents the results from the analysis of race-related words in each subcorpus.

### **6.1. Keywords in the Grime Corpus**

The keyword calculation retrieved nine words in the broadsheets and five words in the tabloids. All keywords related to the theme of music. They are presented in Tables 3 and 4, ordered by their frequency

of occurrence in each subcorpus. In the tabloids, the calculation yielded four words referring to the themes of cleaning ('clean' and 'cleaning') and skincare ('skin' and 'make-up'). These words were discarded given that the present analysis seeks to examine discussions around racial issues in the context of grime music specifically. However, it is important to explain that this was not surprising as it is directly associated with the very nature of language; words have multiple meanings. An inevitable side-effect of an automatic collection of texts is that by using a given word as query term (in our case, 'grime'), the procedures retrieve any text containing that word, irrespective of meaning. A selection of individual texts according to meaning would require reading the retrieved documents, which would go beyond the scope of the project. At the same time, the presence of irrelevant texts in the subcorpus was not viewed as a problem since they had no significant impact on the analysis.

<i>Themes</i>	<i>Keywords</i>	<i>Raw Frequency</i>	<i>Keywords</i>	<i>Raw Frequency</i>
<i>Music</i>	album	977	songs	367
	festival	568	debut	289
	artists	455	jazz	279
	artist	396	track	276
	pop	389		

Table 3: 'Keywords' in the Broadsheet Grime subcorpus, organised by theme

<i>Themes</i>	<i>Keywords</i>	<i>Raw Frequency</i>	<i>Keywords</i>	<i>Raw Frequency</i>
<i>Music</i>	star	106	artist	63
	festival	105	tonight	61
	album	80		

Table 4: 'Keywords' in the Tabloid Grime subcorpus, organised by theme

As shown, the keywords in both the Broadsheet and the Tabloid Grime subcorpora related to music. Mentions of artists were frequent (cf. 'artist(s)' in the broadsheet, and 'star' and 'artist' in the tabloids), with reference to grime artists as well as other types of music such as pop, jazz, hip-hop, drill, and rock. An exploration of the context in which the keyword 'festival' in both subcorpora unveiled frequent mentions of artists who performed in music festivals (such as Glastonbury, Bestival music festival in Dorset, and Leeds festival). The newspapers also mentioned albums (cf. 'album' in both tables), very often with regard to their release; the word 'album' is frequently qualified by words such as 'debut', 'first', 'new' or 'latest'. The discussion in the broadsheet goes further and also highlights specific 'songs' or 'track'.

## 6.2. References to race in the Grime Corpus

We now turn our attention to words which could function as references to race. These words are shown in Figures 9 and 10 in terms of their relative frequency (per 100,000 words) within each component

of the Grime corpus: broadsheet and tabloids. Note that the figures include the occurrences of the selected BAME-related words as well as ‘white’, used here a benchmark for comparison.

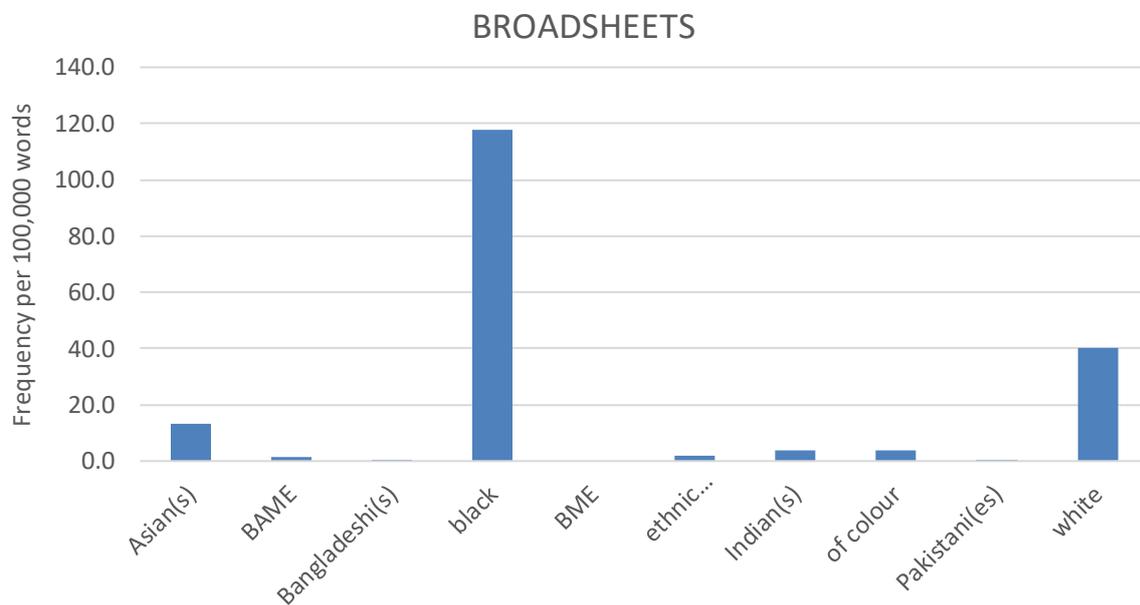


Figure 9: Relative frequency of race-related words in the Broadsheet Grime subcorpus

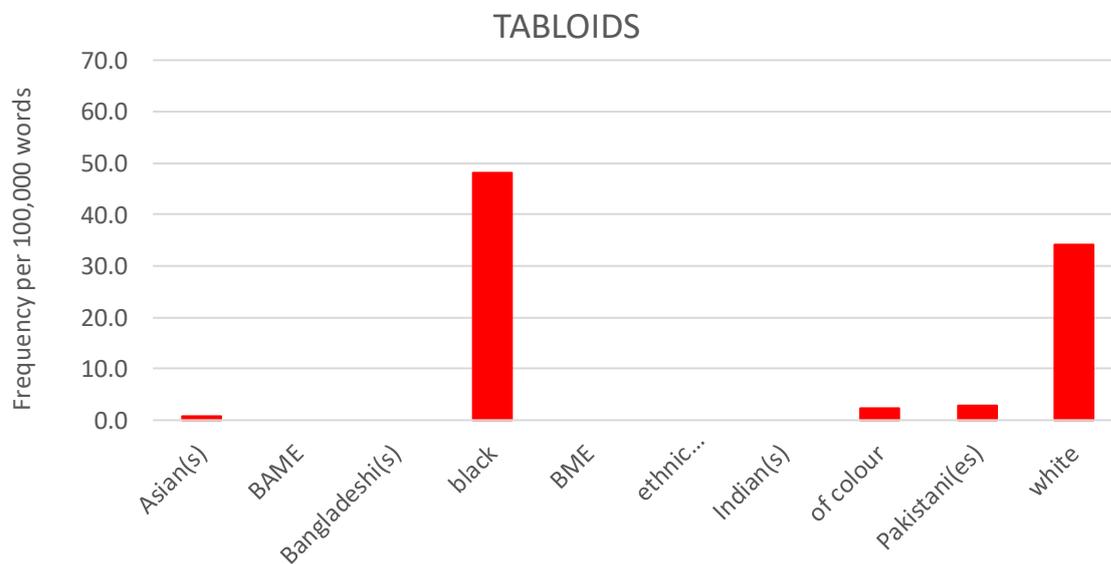


Figure 10: Relative frequency of race-related words in the Tabloid Grime subcorpus

As shown, ‘black’ is the most frequent race-related word in both Grime subcorpora, with 117.6 instances per 100,000 words in the broadsheets and 48.2 in the tabloids (873 and 65 occurrences respectively). In both the broadsheet and the tabloid subcorpora, the word ‘white’ is the second most frequent race-related word after ‘black’: with 40.1 and 34.1 occurrences per 100,000 words in the broadsheets and tabloids respectively (298 and 46 instances). All other words appeared much less frequently.

In the broadsheet subcorpus, the third most frequent word is ‘Asian(s)’, with 13.2 occurrences per 100,000 words (98 instances). In the tabloids, the gap between the second and the third most frequent race-related word is even wider. ‘Pakistani(es)’ is the third most frequent race-related word, with only 3.0 occurrences per 100,000 words (four instances).

We now explore the most dominant discourse patterns around the words ‘black’ and ‘white’, the two most frequent race-related words in both the Broadsheet and the Tabloid Grime subcorpora. This is done by examining their collocates (Figures 11 and 12), which are the words frequently appearing around them (see section 5).

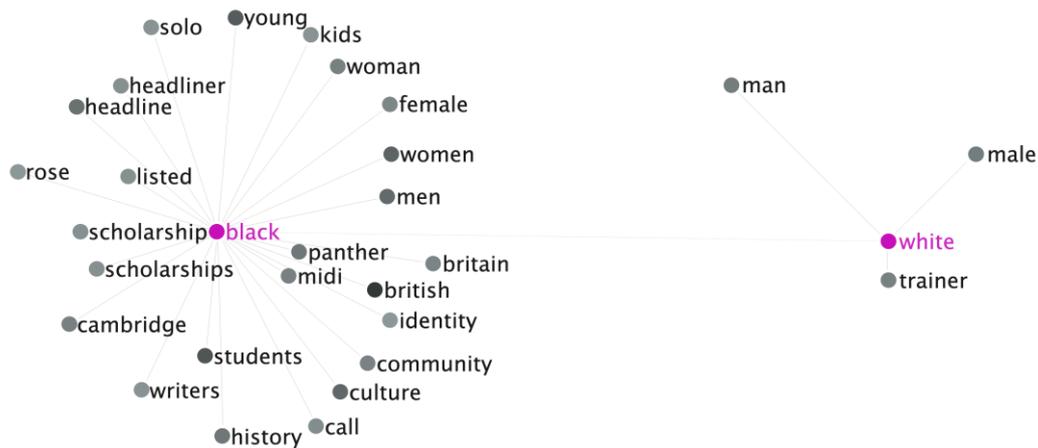


Figure 11: Collocations of ‘black’ and ‘white’ in the Broadsheet Grime subcorpus

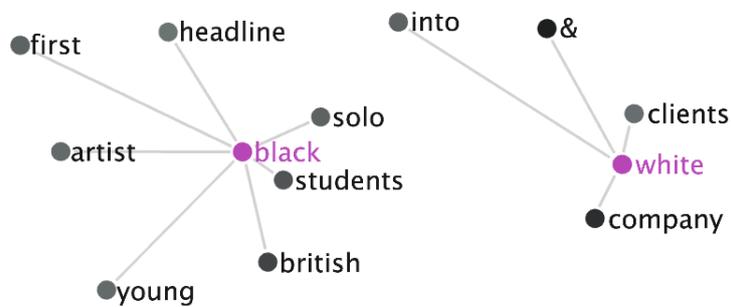


Figure 12: Collocations of ‘black’ and ‘white’ in the Tabloid Grime subcorpus

Examining the collocates of ‘black’, we find that it collocates with the word ‘British’ in both subcorpora (see Figures 11 and 12). In fact, ‘British’ is the most frequent collocate of ‘black’ in both subcorpora: they co-occur 118 times in the broadsheets and 13 times in the tabloids. Most of these instances refer to the phrase ‘black British’, which accounts for 78% of instances in the broadsheets (92 out of 118) and 85% in the tabloids (11 out of 13). Combined with the collocates ‘solo’, ‘first’, ‘artist’, and ‘headline’ in both subcorpora, as well as ‘headliner’ and ‘history’ in the broadsheet subcorpus, the collocations of ‘black’ with ‘British’ uncovered frequent mentions of the British rapper Stormzy being the first black British solo artist to headline Glastonbury (Example 1). The broadsheet newspapers went further and also reported

on Stormzy's initial misstatement that he was the first artist to headline Glastonbury and later apology to Skunk Anansie for the mistake (Example 2).

(1) *Sir Attenborough SWNS Stormzy, who became the first black British solo artist to headline the festival in its 49-year history, yesterday told fans his set had been 'the greatest night of my life'. (Metro, 01.07.2019).*

(2) *The grime star Stormzy has apologised to Skunk Anansie for saying he was the first black British artist to headline Glastonbury. (i-Independent, 01.07.2019)*

Broadsheet newspapers also mentioned Beyoncé and Nelson Mandela in the spotlight at the Glastonbury festival as well as notable achievements of black people across other sectors, such as theatre and literature. This was especially evident in relation to women, as indicated by the collocations of 'black' with 'woman' or with 'female' (Example 3). It is interesting to note that, irrespective of the context, both the broadsheets and the tabloids presented these as exceptional cases, as evidenced through the use of words such as 'first' and 'only'.

(3) *At this year's Edinburgh festival fringe, her [London Hughes'] sold-out, roof-raising show To Catch a Dick narrowly missed out on the big prize, making her the first black British woman to be nominated. (The Guardian, 03.09.2019)*

The broadsheet reporting also includes frequent references to British black culture and identity. This was seen not only through the collocations of 'black' with 'culture' and 'identity' but also through frequent mentions of black British music, artists, talents, and experience. These are illustrated in Figure 13 below, which shows a set of concordance lines (the preceding and following stretches of text) of the phrase 'black British'.

the series could help readers outside **black British** communities understand the impor  
new non-fiction series focused on the **black British** community, all written by young bl  
Labour MP for Tottenham. Black, Listed: **Black British** Culture Explored by Jeffrey Boakye  
official: Stormzy is a phenomenon and **black British** culture is flying; The artist's mi  
ack Britain at the time? The extent of **black British** culture seemed to be one black fam  
t what she does say is: "I wanna bring **black British** culture to America; the positive  
addresses it in a forthcoming book on **black British** culture, Black, Listed. "One simp  
force behind this reboot of the cult **black British** drama starring former So Solid Cr  
le Black, a powerful reflection on the **black British** experience - spanning African dias  
tion books that are grappling with the **black British** experience. It burns with purpose  
ng black authors; A groundbreaking new **black British** history series will explore subjec  
med into a multifaceted celebration of **black British** life, adding musicians, gospel sin  
him, this boom in the popularity of **black British** music culture would not be happeni  
der. Aubrey Drake Graham has championed **black British** music from day one, sampling our  
a good airing. The entire spectrum of **black British** music gets a look-in, though. No  
e hardly surprising. On the face of it, **black British** music is more successful than eve  
minantly white mainstream media of what **black British** music is, and can be. On Basic Vol  
music to the ascension of grime. "Where **black British** music took off around five years  
story of the origins of a contemporary **black British** music whose vitality and urgency  
23) This makes it the first play by a **black British** playwright to open there, and now  
ic to bamboo earrings, plantain to the **black British** power movement, Afrobeats to black  
nther and Moonlight, and a new wave of **black British** talent, from director Steve McQuee  
It proved to be a hothouse of young **black British** talent: Michaela Coel had a bit-p  
with ELLE magazine where he and other **black British** talents were highlighted for their  
think theatre is and how contemporary **black British** theatre smashes these traditional

Figure 13: Instances of 'black British' in the Broadsheet Grime subcorpus

Another relevant aspect of the narrative in the broadsheet newspapers relates to the disadvantageous position of black people in society, and black women in particular, as indicated by the collocations of ‘black’ with ‘women’, ‘woman’ and ‘female’ (Figure 11). Some articles highlighted the absence of black women’s voices in the music world (Example 4).

- (4) *The music industry as a whole has a dearth of black women. Whether it is pop music or more so-called "urban" sounds such as UK rap and Afro-bashment, black female artists are very rarely heard above the underground. (The Guardian, 01.04.2019)*

Others focused on the challenges faced by black women in life, making it explicit that they have to work harder than anyone else. This was sometimes voiced by black women themselves, as in the extract below (Example 5), taken from an interview with Chelsea Kwakye about her book entitled *Taking Up Space: The Black Girl's Manifesto for Change* which she co-authored with Ore Ogunbiyi.

- (5) *There's a common belief among black women – conscious and unconscious – that we need to work twice as hard to get half as far, or [be as] good. I was always a firm believer in working as hard as possible just so I could be on par with average or mediocre people. I know that my race and gender disadvantage me before I even open my mouth. (The Observer, 15.07.2019)*

While a few articles mentioned the association of black men with crime, the negative representation of black men was in many cases counter-balanced with references to their artistic creativity. This was seen, for example, in the broadsheet reporting of the film director Shola Amoo’s intentions in *The Last Tree* (Example 6).

- (6) *"We know young black men are routinely criminalised, but I wanted to let black kids have moments of pure teenage joy that have always been denied them on screen. ..." (The Guardian, 19.09.2019)*

The word ‘black’ was also found to co-occur with ‘students’ in both subcorpora (see Figures 11 and 12). In the broadsheet subcorpus, it also collocated with other words related to education (cf. ‘scholarship(s)’, ‘Cambridge’ and ‘writers’ in Figure 12). These collocations unveiled several mentions of an initiative by the British rapper Stormzy to support young black people in higher education. The newspapers specifically mentioned his decision to sponsor two black students at Cambridge University, as a response to the very low admission rates of black students by Cambridge colleges (Example 7). The broadsheet newspapers additionally reported on Stormzy’s launch of a publishing house for black writers specifically as a way to encourage them to embark in writing (Example 8).

- (7) *The rapper launched a scholarship after learning that no black students got into some Cambridge colleges from 2012 to 2016. (The Sun, 10.11.2018)*
- (8) *Stormzy, who has funded scholarships at Cambridge University for black students, will launch #Merky Books, a Penguin imprint, next month to find new, mainly young and black, writers. (The Sunday Times, 28.10.2018)*

As for ‘white’, it co-occurs with ‘black’ 38 times in the broadsheet subcorpus (Figure 11). In the tabloids, there is no association between the two words (Figure 12). An exploration of their collocations in the broadsheet newspapers showed that, when co-occurring, these two words mostly refer to race. About a quarter of those 38 instances (24%) came from one single article, published by *The Guardian* and written by a black writer (Jeffrey Boakye) who has a white wife. However, most of the discussion around race relies on the binary racial definition that distinguishes the ‘black’ from the ‘white’ race. While a few instances

reflected a positive interaction between the two (Example 9), most instances exposed differences in treatment across various settings, and in the music world in particular, sometimes challenging traditional stereotypes (Example 10).

(9) *Piers Aggett, who is one quarter of the dance music group Rudimental whose members are white, black, mixed-race and Middle Eastern, said it was Britain's diverse races and cultures that had made the country's popular music scene so rich. (The Guardian, 15.03.2019)*

(10) *Gaika pays no attention to the limitations of traditional genre or song writing structure, tackling themes of identity and the tensions of life in the city with profound intelligence; consistently challenging tired stereotypes put in place by a predominantly white mainstream media of what black British music is, and can be. (The Independent, 03.12.2018)*

Interestingly, the collocates of ‘white’ with ‘man’ and ‘male’ reinforce this last trend by uncovering mentions of biases towards white people, and white males in particular (Example 11).

(11) *Diversity of titles means music journalism is more widespread, but it's still predominantly white, straight and male. (The Guardian, 24.10.2018)*

In the tabloids, the word ‘white’ was rarely employed to refer to race, thus of no relevance for the context of this study. As its collocates indicate, over a third of the occurrences of ‘white’ (35%) appeared in one single article published by the *Daily Mail*, which reported on a fraud investigation of an accountancy firm name *White & Company*. The company shared offices with an event organisation enterprise that ran a festival in which the grime rapper Dizzee Rascal featured.

## 7. The Stop-and-Search corpus

This section focuses on the analysis of the Stop-and-Search Corpus. Following the same structure used in the previous section, it is divided into two sub-sections. The first sub-section discusses the keywords in the Broadsheet and Tabloid Stop-and-Search subcorpora and the second sub-section presents the results from the analysis of race-related words in each subcorpus.

### 7.1. Keywords in the Stop-and-Search Corpus

The keyword calculation generated 40 words in the broadsheets and 52 words in the tabloids. Tables 5 and 6 show the keywords in each subcorpus, grouped thematically and ordered by frequency of occurrence.

<i>Themes</i>	<i>Keywords</i>	<i>Raw Frequency</i>	<i>Keywords</i>	<i>Raw Frequency</i>
<i>Crime</i>	crime	2082	drugs	259
	knife	1289	knives	242
	attacks	294	crimes	212
	stabbed	277	weapons	205

<i>Offenders</i>	youth	306	gangs	221
	gang	270		
<i>Law Enforcement</i>	police	2685	policing	432
	stop	1406	police officers	303
	search	1162	prison	267
	officers	845	tackle	251
	powers	525	commissioner	185
<i>Violence</i>	violence	940	violent	506
<i>Politics</i>	deal	670	leadership	329
	EU	598	speech	273
	Prime Minister	502	parliament	268
	tory	479	conservative	259
	home secretary	419	home office	258
	MPs	388	mayor	191
	Hunt	362	Commons	186
	secretary	338	cabinet	184
<i>Finance</i>	cuts	359		

Table 5: ‘Keywords’ in the Broadsheet Stop-and-Search subcorpus, organised by theme

<i>Themes</i>	<i>Keywords</i>	<i>Raw Frequency</i>	<i>Keywords</i>	<i>Raw Frequency</i>
<i>Crime</i>	crime	1,117	murder	114
	knife	1,054	to death	110
	knives	234	drugs	95
	stabbed	205	attacks	81
	crimes	131	offences	69
	carrying	124	blade	58
	weapons	122		
<i>Offenders</i>	criminals	147	youth	79
	gang	134	thugs	72
	offenders	128	teenagers	66
	gangs	125		

<i>Law Enforcement</i>	police	1211	prison	128
	stop	667	police officers	124
	search	479	sentences	100
	officers	343	jail	76
	powers	217	chiefs	73
	policing	147	cops	73
	tackle	136	commissioner	69
<i>Violence</i>	violence	260	violent	212
<i>Places</i>	streets	221	on the streets	89
	West Midlands	60		
<i>Politics</i>	home secretary	220	home office	75
	prime minister	150	politicians	74
	Tory	120	leadership	69
	MPs	97	mayor	66
<i>Threat</i>	to death	110	emergency	68
	crisis	106	fatal	58
	epidemic	100		
<i>Finance</i>	cuts	87		

Table 6: 'Keywords' in the Tabloid Stop-and-Search subcorpus, organised by theme

Not surprisingly, most keywords in both the broadsheet and the tabloid Stop-and-Search subcorpora fell under the themes of crime and law enforcement, and both 'violence' and 'violent' were found among the top keywords. In both subcorpora, much of the discussion revolves around crimes, and knife crimes in particular, as indicated by the presence of keywords such as 'knife/knives' and 'stabbed' (Example 12-13). The words 'youth' in the broadsheets and 'teenagers' in the tabloids indicates that knife crime and the stop-and-search approach is mostly associated with young people.

(12) *As MPs and police chiefs continued to row over how to resolve the crisis, the Daily Mail approached the families of some of the 27 teenagers stabbed to death in the last year to hear their views.* (The Daily Mail, 27.03.2019)

(13) *'Like fires everywhere': West Midlands becomes a youth knife-crime hotspot; The causes are numerous and complex, but it is the scale and the age of those involved that alarms adults on the frontline.* (The Guardian, 29.11,2018)

There is an emphasis on policies that can strengthen the police's powers to stop, search and detain suspects (Example 14-15). The discussion of policies also explains the salience of political references, especially in relation to a significant decrease in financial resources made available to police and youth services (cf. 'cuts').

(14) *And it was last night revealed he [the home secretary] wants to give cops increased stop and search powers to combat knife crime.* (The Sun, 09.11.2018)

(15) *The Government is working on a Drones Bill, which will give police powers to stop and search people suspected of using drones maliciously above 400ft or within three miles of an airport.* (The Daily Telegraph, 20.02.2019)

## 7.2. References to race

We now turn our attention to words which could function as references to race in the Stop-and-Search corpus. These words are shown in Figures 14 and 15 in terms of their relative frequency (per 100,000 words) within each corpus. The figures also include the occurrences of the word ‘white’, as a benchmark for comparison.

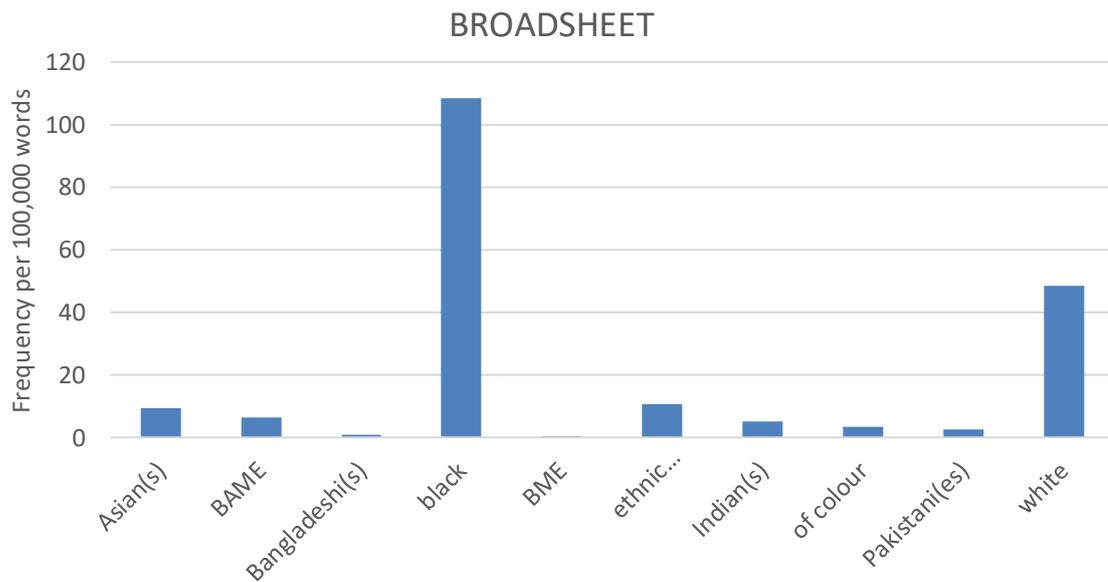


Figure 14: Relative frequency of race-related words in the Broadsheet Stop-and-Search subcorpus

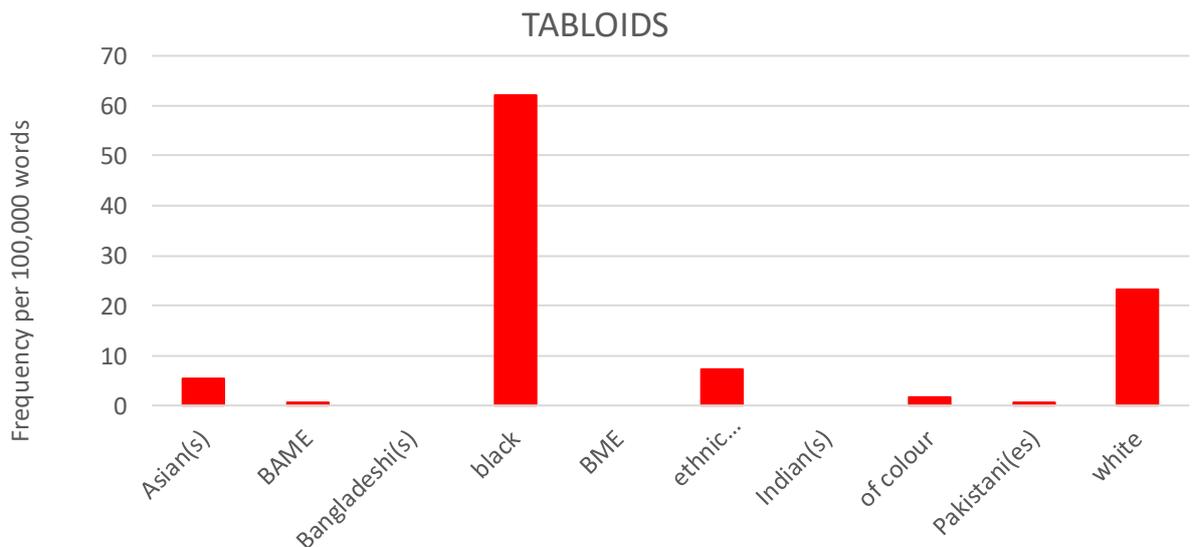


Figure 15: Relative frequency of race-related words in the Tabloid Stop-and-Search subcorpus

As shown, ‘black’ is by far the most frequent race-related word in both Stop-and-Search subcorpora (108 occurrences per 100,000 words in the broadsheet and 62 occurrences in the tabloids), appearing ten

times more frequently than the second most frequent BAME-related word ('ethnic minority(ies)': 11 and 7 occurrences per 100,000 words in the broadsheet and tabloids respectively). Interestingly, the word 'white' is the second most frequent race-related word after 'black' in both subcorpora, with about half of the number of instances of 'black': 48 and 23 occurrences per 100,000 words in the broadsheet and tabloids respectively. As we shall see later, this mainly relates to the fact that 'white' is used in contrast with 'black' in relation to how racial issues operate in the context of the stop-and-search measure.

We now turn our attention to the word 'black' and examine the contexts in which it is used by looking at its collocates in the Broadsheet and Tabloid Stop-and-Search subcorpora (Figures 16 and 17 respectively).

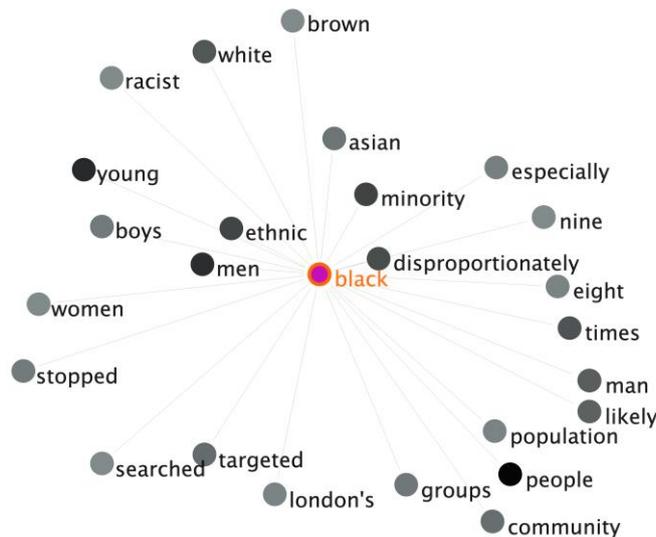


Figure 16: Collocations of 'black' in the Broadsheet Stop-and-Search subcorpus

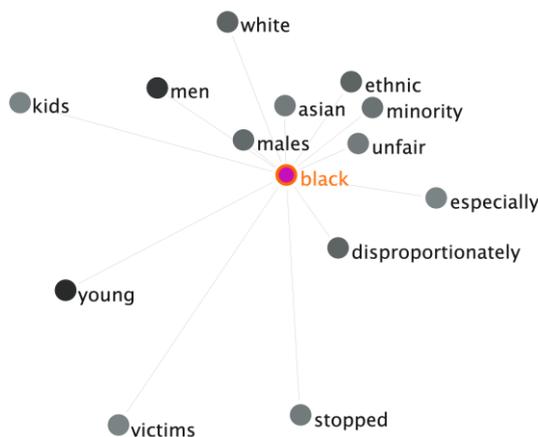


Figure 17: Collocations of 'black' in the Tabloid Stop-and-Search subcorpus

The reporting in both the broadsheets and the tabloids expose bias against black people, and black men in particular, by highlighting their stereotypical association with crime. Through the collocations with 'disproportionally', 'likely', 'stopped', 'especially', and 'searched', the newspapers stress that black people are much more likely to be stopped and searched than white people are (Example 16). Some broadsheet

articles provide the proportion in which this occurs by stating that the probability is ‘eight’ or ‘nine’ ‘times’ higher (Example 17).

(16) *But stop and search disproportionately targets black people and there's no evidence that it actually reduces crime.* (The Mirror, 14.08.2019)

(17) *Stop and search is 'ineffectual and racially unjust', says David Lammy. Police 'nine times as likely' to stop and search black people.* (The Independent, 25.11.2018)

This stereotypical representation is especially evident in relation to black males, and young black males in particular (cf. ‘men’, ‘males’ as well as ‘young’, ‘boys’ and ‘kids’). In both the broadsheet and the tabloids, we found mentions of black males not only as victims of the system but also as perpetrators (Example 18-19). What is more, racial injustice is not presented as being limited to knife crimes specifically. We found evidence that unfairness towards black people, and to a lesser extent towards Asians and other minority groups, is presented as being more widely spread in the system, with regard to crimes in general and legal procedures (Example 20).

(18) *If you also look into lives being lost- particularly in London- they are predominantly not white. In recent years, the number of black murder victims has been four times higher when compared with white victims.* (The Express, 27.03.2019)

(19) *Unfortunately, the issue of knife crime in England is tied up with race. In Scotland, the street gangs are white, while London gangs are multicultural. Black and minority ethnic teenage boys are disproportionately affected by knife crime in the capital, both as victims and perpetrators.* (The Sunday Times, 28.07.2019)

(20) *When it comes to drugs, black people are nine times more likely to be stopped and searched. And if they are found to possess illicit substances, they suffer more punitive treatment.* (The Guardian, 19.10.2018)

Thus, the analysis of the collocations of ‘black’ uncovered several mentions that black and white people are not treated equally in relation to crime and justice. Further evidence of reports on disparities in treatment came from the collocations of the word ‘white’ in the two corpora (Figures 18 and 19).

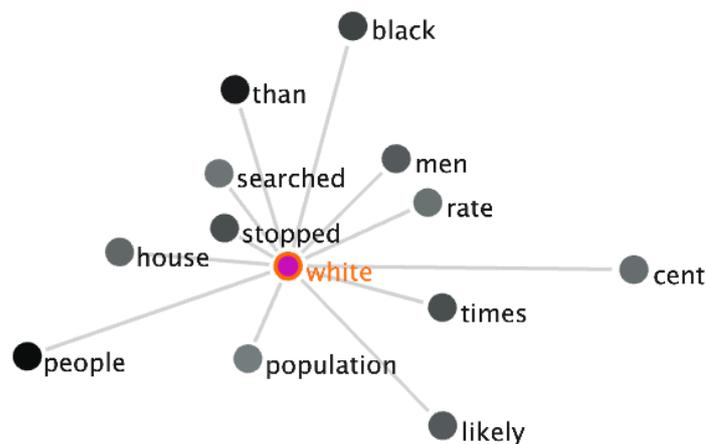


Figure 18: Collocations of ‘white’ in the Broadsheet Stop-and-Search subcorpus

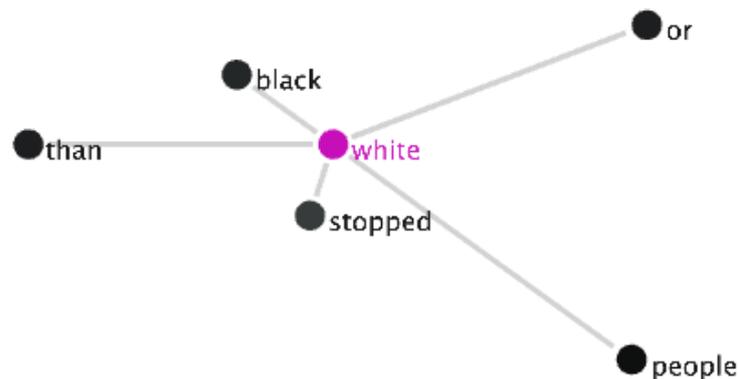


Figure 19: Collocations of ‘white’ in the Tabloid Stop-and-Search subcorpus

As in the case of ‘black’, the collocations of ‘white’ also point towards mentions of the frequent comparisons between the ways black and white people tend to be treated. As indicated by the collocations with, for example, ‘black’ and ‘than’, the word ‘white’ is mainly used to expose the bias against black people. The broadsheets specifically stress the rate at which the disparity occurs, thus making it explicit that there is no racial justice in the system. This is evident in *The Observer*’s reporting of a study by the London School of Economics and Political Science, a coalition of academics, lawyers, civil society organisations (Stopwatch) and a centre of expertise on drugs and drugs law (Release) (Example 21). In the tabloids, the collocations of ‘white’ with ‘or’ uncovered mentions that the stop-and-search policy should be applied to anyone, irrespective of race; an argument that was sometimes used to reiterate the claim that the police powers should be strengthened (Example 22).

(21) *It also reveals that black people made up a quarter of those convicted of cannabis possession, despite making up less than 4% of the population. The proportion is even more striking as it is considered that black people report using cannabis at slightly more than half the rate of white people.* (The Observer, 13.10.2018)

(22) *Whether you're Pink, Green, Black or White, it's time the Police had more power to stop and search. If people have nothing to hide, then what's the problem?* (The Daily Star, 08.03.2019)

## 8. The Race Equality corpus

In this section, we present the findings of the analysis of the Race Equality Corpus. It follows the same structure as that used in the previous two sections. The first sub-section discusses the keywords in the two subcorpora and the second sub-section presents the results from the analysis of race-related words in each.

### 8.1. Keywords in the Race Equality Corpus

The keyword calculation retrieved 29 words in the broadsheets and 25 words in the tabloids. Tables 7 and 8 show the keywords from each subcorpus, grouped thematically, and ordered by frequency of occurrence.

<i>Themes</i>	<i>Keywords</i>	<i>Raw Frequency</i>	<i>Keywords</i>	<i>Raw Frequency</i>
<i>Race</i>	racial	822	race	500
	racism	528	racist	352
<i>Justice</i>	equality	394	inequality	185
	justice	344	injustice	183
	diversity	186		
<i>Crime</i>	racism	528	discrimination	204
<i>BAME people</i>	minority	185	black	1,352
<i>Demonstration</i>	protest	191	march	198
<i>The U.S. Elections</i>	president	650	democrats	335
	campaign	472	America	305
	democratic	448	voters	262
	candidates	383	Americans	203
	debate	359	candidate	195
<i>The U.S. Football League</i>	super	303	quarter	219
	bowl	272		
<i>Newspaper</i>	Guardian	178		

Table 7: 'Keywords' in the Broadsheet Race Equality subcorpus, organised by theme

<i>Themes</i>	<i>Keywords</i>	<i>Raw Frequency</i>	<i>Keywords</i>	<i>Raw Frequency</i>
<i>Race</i>	racial	60	racist	20
	race	29		
<i>Justice</i>	equality	39	inequality	10
	injustice	13		
<i>BAME people</i>	black	79	Asian	12
<i>Crime</i>	racism	24	incidents	9
	discrimination	10	rape	9
<i>Art</i>	Shakespeare	10	movie	12
	film	30		
<i>Sport</i>	football	22	Smith	11
<i>Activists</i>	King (Martin Luther King)	51		

<i>Royal family</i>	queen	15	royal	11
<i>Politics</i>	leader	21	Tory	14
<i>Miscellaneous</i>	nation	16	legacy	9
	jokes	9		

Table 8: ‘Keywords’ in the Tabloid Race Equality subcorpus, organised by theme

Not surprisingly, racial issues emerged as dominant themes in both the broadsheet and tabloid newspapers, as indicated by the combination of the keywords related to race and justice. These were especially discussed in relation to discrimination (cf. ‘discrimination’ and ‘racism’) against BAME people, and black people in particular. Note that the keyword ‘black’ has the highest frequency of all keywords in both subcorpora.

Three additional themes emerged as salient in the reporting of broadsheet newspapers (Table 7). One issue related to protests in the UK and the US, mostly those against racial injustice and violence against women. The other two themes referred to the US context specifically: the American elections and the American Football league. The word ‘Guardian’ emerged as a keyword because many articles published by *The Guardian* – which accounted for more than half of the articles in the broadsheet subcorpus (55%, see Figure 5) – made reference to other article previously published by the newspaper.

As for the tabloids (Table 8), there were also mentions of ‘rape’ crimes but most instances referred to articles about films and books which had rape as part of the plot. In fact, many articles were reviews of films, Shakespeare’ plays, some of books (cf. keywords under the theme of art). There were also those which mentioned racial issues in sports, and in ‘football’ in particular. Mentions of ‘Smith’ were mostly to do with the US gold-medal Tommie Smith in an article by *The Express*, which referred to his gesture of bowing his head and raising his black-gloved fists in the 1968 Olympic Games in Mexico City to represent black power in America. Rather than about activism, references to Martin Luther King related to an article in *The Daily Mail* about a FBI report stating that Martin Luther King had watched one of his assistants commit a rape crime. All mentions of the royal family came from one single text published by *The Express*.

## 8.2. References to race

Figures 20 and 21 show the relative frequency (per 100,000 words) of race-related words in the Broadsheet and the Tabloid Race Equality subcorpora.

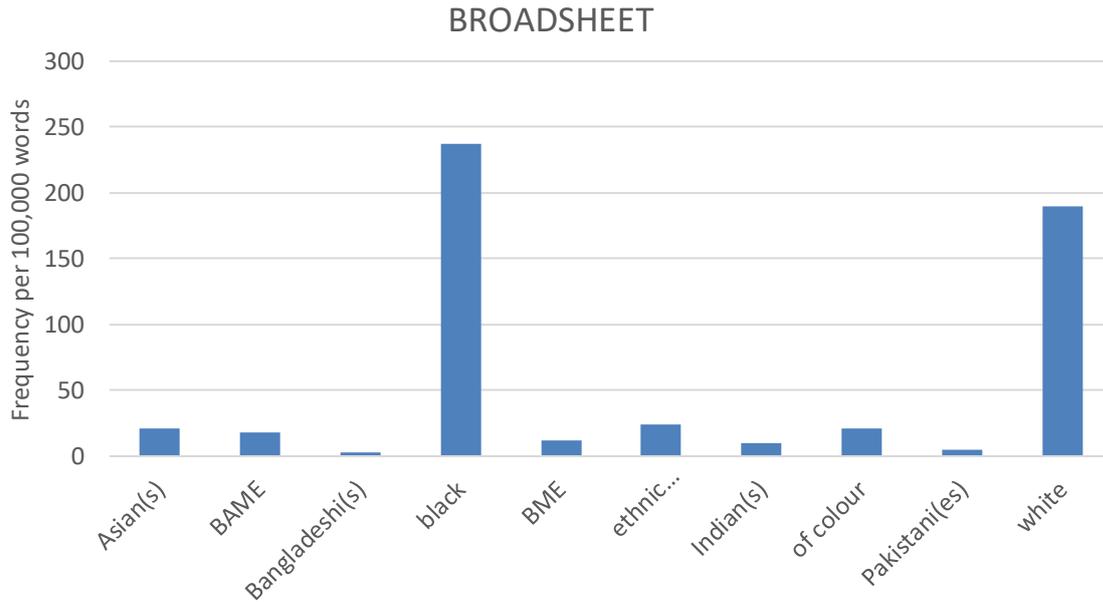


Figure 20: Relative frequency of race-related words in the Broadsheet Race-Equality subcorpus

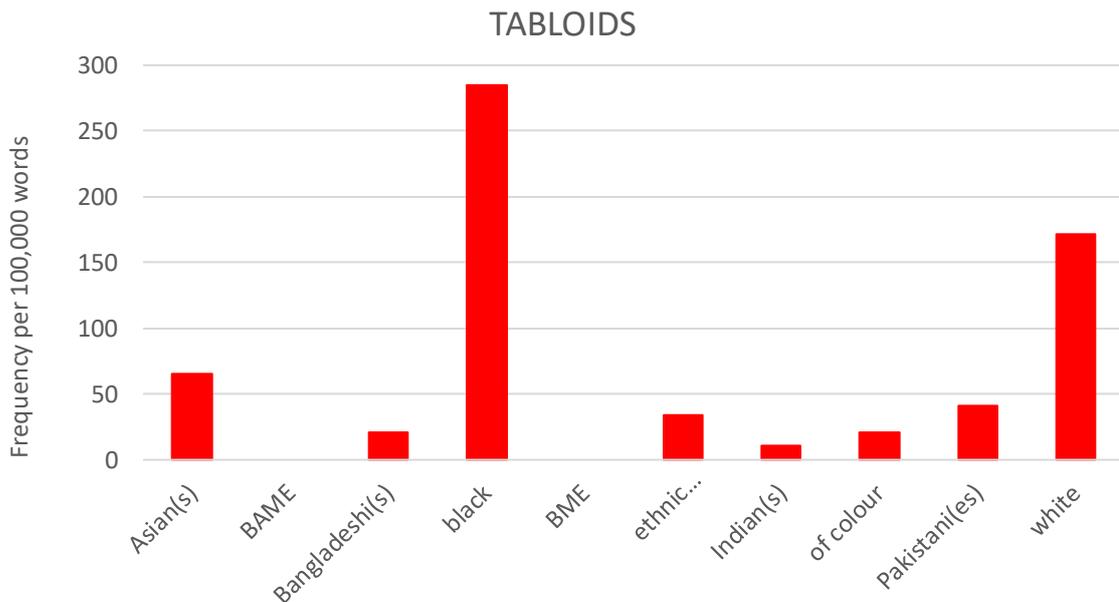


Figure 21: Relative frequency of race-related words in the Tabloid Race-Equality subcorpus

Following the trend seen in the previous two corpora, here again the word ‘black’ is the most frequent race-related word in both subcorpora: 237.2 occurrences per 100,000 words in the broadsheet (1,405 instances) and 285.3 occurrences in the tabloids (83 instances). Within BAME-related word in the broadsheet subcorpus, ‘black’ is ten times more frequent than the second most frequent word (‘ethnic minority(ies)’): 23.6 occurrences per 100,000 words (140 instances). In the tabloid subcorpus, ‘Asian(s)’ is the second most frequent BAME-related word, with 65.3 occurrences per 100,000 words (19 instances). In both subcorpora, ‘white’ is the second most frequent race-related word after ‘black’, with 189.4 and 171.8 occurrences per 100,000 words respectively (1,122 and 50 instances).



men or white women (Example 24). A similar picture emerges when we examine the collocations of ‘white’ with ‘men’, ‘man’ and ‘male’. The reporting highlighted privileges of white males in society, as compared to other races, and black men in particular (Example 25). References to white supremacy (cf. ‘supremacist(s)’, ‘supremacy’ and ‘nationalists’ relate to the US context specifically.

(23) *In the wake of gender pay gap reporting laws, the government will consult over forcing thousands of businesses to disclose disparities between the pay packets of black, Asian and ethnic minority (Bame) employees and their white counterparts. (The Independent 11.10.2018)*

(24) *In interviews with 20 of the total 25 black female professors working in UK universities, Nicola Rollock, the report's author, said that their experiences made for depressing reading. "What they are saying is that their entire careers have been characterised by abuse and exclusion, and that their race has been the key to that," said Rollock. (The Guardian, 04.02.2019)*

(25) *As a number of influential figures (themselves from ethnic minority backgrounds) wrote to this paper in response to the Race Disparity Audit, "All too often statistics are misused in a way that casts minorities as victims of racism and 'white privilege'". (The Times, 15.10.2018)*

In the tabloids (Figure 23), the collocations of ‘black’ with ‘history’ pointed towards mentions of ‘Black History Month’, with all instances being found in articles published in October 2018. The tabloid reporting revolved around a statement by the Labour leader (Jeremy Corbyn) at the time, who suggested that future generations should be taught about the contribution of the Black community to British society and the legacy of British colonisation. Articles in *The Daily Mail* fiercely criticised Corbyn’s position (Example 26). As in the case of the Stop-and-Search corpus, here again the tabloids exposed the bias against black males (cf. ‘man’), and young black males in particular (cf. ‘young’). This was especially discussed in relation to interaction between black men and the police, sometimes voiced by a black male himself (Example 27). The collocations of ‘white’ in the tabloids uncovered mentions of racial disparities to the disadvantage of BAME people across various settings, including work, sport and education (Example 28).

(26) *But to Mr Corbyn, it's not enough that every October is set aside as black history month. He wants the subject taught the whole year round, saying: 'Black history is British history. It is vital that future generations understand the role black Britons have played in our country's history and the struggle for racial equality.' (The Daily Mail, 12.10.2018)*

(27) *Not surprising; I was just eight years old and the sight of two policemen beating a black man to a bloody pulp seemed like a clear warning: get used to the treatment that awaits you when you grow up. (The Daily Mail, 23.02.2019)*

(28) *BOSSES will be made to reveal salary differences between their white and ethnic minority staff. (The Daily Mail, 11.10.2018)*

## 9. Conclusion

Overall, the analysis demonstrates that racial issues emerged in the reporting of both the broadsheets and the tabloids across the three corpora. There are several mentions in the reports that BAME people are

unfairly treated, and black people in particular, regardless of gender. More specifically, we can summarise as follows how the UK press represents BAME people in the three contexts analysed here.

- Highly successful black artists in the music world are presented as exceptional cases, and black women in particular.
- The broadsheet reporting seems to expose a bias towards white artists, and white males in particular. Several articles mention differences in the ways black and white artists are treated in the music world, highlighting it is predominantly white and male.
- The reporting in both the broadsheets and the tabloids highlight the stereotypical association of black males with crime, and young black males in particular, by frequently mentioning that black males are more likely to be stopped and searched.
- In both the broadsheet and the tabloids, we found mentions of unfairness towards black people, and to a lesser extent towards Asians and other minority groups, as being widely spread in the system, with regard to crimes in general and legal procedures.
- Both the broadsheets and the tabloids highlight that black people, regardless of gender, are more likely to be victims of crimes.
- Racial disparities to the disadvantage of BAME people were frequently mentioned in the newspapers, and in the broadsheet in particular. This was reported to occur across various settings, such as work, education, sports and society at large.
- Broadsheet newspaper specifically mentioned that black women face even greater challenges than black men or white women do.

## References

- Brezina, V., McEnery, T. & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173, [doi:10.1075/ijcl.20.2.01bre](https://doi.org/10.1075/ijcl.20.2.01bre).
- Hardie, A. (*forthcoming*). A dual sort-and-filter strategy for statistical analysis of collocation, keywords, and lockwords. An informal introduction can be found at <http://cass.lancs.ac.uk/?p=1133>; accessed 2020-04-01.
- McEnery, T., Hardie, A. (2011) *Corpus linguistics : method, theory and practice*. Cambridge: Cambridge University Press.
- Rayson, P. (2009). *Wmatrix: a web-based corpus processing environment*. Lancaster, Computing Department, Lancaster University.