



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Triennale Interclasse in
Lingue, Letterature e Mediazione culturale (LTLLM)

Classe LT-12

Tesina di Laurea

*Discrimination and racism in language: a
corpus linguistics analysis of media coverage of
Black Lives Matter*

Relatrice

Prof.ssa Fiona Dalziel

Laureanda

Virginia Bonollo

n° matricola 1229939 / LTLT

Anno Accademico 2021 / 2022

'A riot is the language of the unheard'
Martin Luther King

Table of contents

Abstract.....	7
Introduction.....	9
Chapter 1: How minority representation can be influenced.....	11
1.1 The role of discourse.....	12
1.2 The role of the media in the reproduction of racism.....	13
1.3 Biased news.....	16
1.4 Discrimination against non-white journalists: a connection to Frederick Douglass.....	17
1.5 The power of headlines: a case study.....	19
Chapter 2: corpus linguistics in discourse analysis.....	23
2.1 Corpus linguistics.....	23
2.1.1 Corpus building.....	25
2.1.2 Corpus size and annotation.....	26
2.2 Corpus-based and corpus-driven approaches.....	27
2.3 Frequency, Concordances and Collocates.....	28
2.4 Using a reference corpus.....	30
Chapter 3: Corpus linguistics analysis of Black Lives Matter	33
3.1 Word frequency.....	34
3.2 Collocations and concordances.....	38
3.3 Connotations, metaphors and hyperbole.....	47
Conclusion.....	51
Bibliography.....	53
Riassunto in italiano.....	57

Abstract

Language is the strongest tool that human beings have as it allows them to express themselves and thus be free. However, in today's society, racism and discrimination are still, unfortunately, very frequent phenomena that we hear about every day in the media. This dissertation, therefore, aims to shed light on the issue of discrimination against minorities from a linguistic point of view. It explores the role of the media and the press in the dissemination of this phenomenon and the various techniques that are used to make the news biased. In the second and third chapters, this paper takes an increasing turn towards the theoretical aspect of corpus linguistics, exploring the advantages of using a corpus for linguistic analysis and the various techniques of analysis. The dissertation ends, in the last chapter, with a real analysis: a special corpus has been built up, composed of articles about the Black Lives Matter activist movement, with which a linguistic study is carried out in order to confirm or refute the thesis put forward in the first chapter, that the depiction of minorities is always exposed in a negative light.

Introduction

The world is making progress in many areas, from technology to sustainability. However, when it comes to racial discrimination, we cannot say the same. We live in an era where 'diversity', be it ethnic, racial, gender, or sexual orientation, is still seen as a threat, a problem to be solved, when in fact it is diversity itself that makes the cities, cultures, environments and spaces we live in, rich. Since language is the tool that allows us to speak, express ourselves, and make us feel free, I wanted to dedicate my dissertation to the issue of racial discrimination, dealt with from a linguistic perspective. The geographical area of interest is English-speaking, and in fact the final analysis of the dissertation focuses on the activist movement for the rights of African-Americans: Black Lives Matter. The pivotal aim of my dissertation is to bring to light how the British and American media report news and facts regarding the struggles for racial equality and to highlight the results through linguistic analysis.

My dissertation is made up of three chapters. The order in which the topics are placed is designed to first provide the reader with general knowledge and concepts regarding the influence of the media and how a linguistic analysis is carried out, and then end with the actual analysis. The first chapter aims to provide an overview of the role played by the media and discourses in the representation of ethnic minorities. The second chapter, on the other hand, leans more towards the main aspects of carrying out a corpus linguistics analysis. The third chapter is the most important because all the concepts mentioned in the first two chapters are included here. In the last chapter, I will carry out the analysis using the analysis tool AntConc.

The first chapter begins with the more general context of what discourses are and how they contribute to the spread of racism (e.g. everyday conversations, parliamentary debates, political propaganda). Then we also look at the role played by the media in the reproduction of racism, which in the 21st century is the main source of information and dissemination. We will see how the language choices and techniques used are able to

shape the interpretation of what is described. Another important feature covered is the discrimination against black journalists and how difficult it is for them to make their voices heard in a world where white supremacy prevails. The last section refers to the role that headlines play in attracting the reader's attention.

The second chapter dives deeper into the theoretical aspect of corpus linguistics analysis. An introduction is given of what a corpus is, how it is built and what the ideal size is. Towards the end of the second chapter, there is a gradual approach to the third chapter because the aspects of frequency, concordance and collocation, which make up the central part of linguistic analysis in the third chapter, are explained.

The third chapter is dedicated to the analysis of a corpus and several sub-corpora created by me, composed entirely of articles concerning the Black Lives Matter movement. I will focus on terms that have a clear connection to racial discrimination, in order to see the contexts in which they are used. The analysis will focus on frequency, concordance, collocation and the different connotations that specific terms can have.

1. How minority representation can be influenced

This first chapter is divided into five sub-sections with the main aim of highlighting how ethnic minorities are described and may be said to be discriminated against in the European press. I will deal with the role played by the media, together with the influence of public discourses. In the fourth section I deal with the topic of discrimination against non-white journalists, with reference to the slave narrative of the 19th century. Finally, the last section analyses a case study on the power that headlines can have over the reader's mind.

The representation of ethnic minorities in the press has to be analyzed from various points of view. It is very important to take into account: the political orientation of the newspaper or website being analyzed; how the news is perceived by readers and especially how and whether it is memorized by them; various linguistic structures; discourse functions and many other factors. I will explore whether the press, might contribute to the spread of racism. I will take into account the percentage of white journalists compared to that of black journalists, especially in the USA; the claims that published news may be biased; the representation of minorities in negative terms or through metaphors or hyperboles.

A definition to summarize this concept, which is quoted in various academic articles, is that of "positive self-representation" and "negative Other-representation". I will investigate claims that minorities are represented as criminals, or as a threat, while those of the majority are the victims or the people who put an end to crime. "Prejudice and racism are not innate but learned" (Van Dijk, 2006:15). These concepts may be assimilated through the press, the media, television news, public discourses and books. Another noteworthy aspect is the role played by the symbolic elites and their influence on the thoughts and ideologies of the readers. These elites (politicians, important people, newspaper editors, activists) have an enormous responsibility: "They control public communication and thus influence, discursively, the beliefs of the dominant group and

its members" (Van Dijk, 2015: 385). However, the elites are the first to deny their involvement in racism. "A country is as racist as its dominant elites are" (Van Dijk, 1991: 6).

1.1 The role of discourse

The simplest and purest definition of discourse is "some form of language use, such as speaking or writing" (Taylor, 2014). Cook (1989: 6-7) provides another definition: "discourse as language in use or language used to communicate something felt to be coherent". But how can this form of language be analyzed? The means is discourse analysis. As reported by Harris (1952) discourse analysis is "a method for the analysis of connected speech or writing, for continuing descriptive linguistics beyond the limit of a simple sentence at a time". As claimed by Cook instead discourse analysis can be described as "the search for what gives discourse coherence [...] what matters is [...] the fact that it communicates and is recognized by its receivers as coherent" (Cook, 1989: 6-7).

According to many racial events and media articles the racism of the 21st century is very different from that of the past¹. The "new racism" of the media is often imperceptible, like a transparent veil that covers and obscures the voice of the discriminated and oppressed. This new racism is expressed mainly through discourses; "they are expressed, enacted and confirmed by text and talk, such as everyday conversations, board meetings, [...] parliamentary debates, political propaganda [...]". (Van Dijk, n.d: 34). In this way, discrimination is expressed in an implicit but hidden way. It can be argued that one tries to use these means in order not to appear too explicit and direct, but on the contrary to always appear on the side of the right and politically correct.

Much of the ideology, prejudice and beliefs against minorities comes from discourse, which tries to influence the minds and thoughts of its readers. When analyzing discourse it is important to define the following: its genre (which may belong to media, political,

¹ *Modern Racism* (<https://www.opendemocracy.net/en/can-europe-make-it/modern-racism/>)

legal, or educational discourse); the topics which are "semantic macrostructures [...] that control the lower-level meaning of the sentences of the discourse" (Zapata-Barrero, Yalaz, 2018: 233) and are the part that is usually remembered most by the reader; it is the overall superstructure which "characterizes types of discourse structure or whole genres - more or less independently of their meaning" (Zapata-Barrero, Yalaz, 2018: 236). In the end, it is essential to define the role and description of actors.

1.2 The role of the media in the reproduction of racism

In the 21st century, the use of the media (such as social media, newspapers, online news and Tv programs) has become widespread and they are now one of the main sources of information and knowledge: "discourse the media produces often functions as the foundation on which people base their conceptualization of the world" (Busà, 2013: 139). Having such an important role in our lives, one might expect them to exert a great influence on our thinking and beliefs. For example very often, the media represent social stereotypes that "contribute to perpetuating specific ideas about social groups" (Busà, 2013: 139).

In support of this statement, I would like to quote an article from *The Guardian* (Rozado, Al-Gharbi, Halberstadt, 2022) which highlights how the use of racist terms has increased dramatically in the American and British press since 1970. Rozado, Al-Gharbi and Halberstadt, analyzed 27 million articles published between 1970 and 2019. The graph below shows the frequency of the use of discriminatory terms in the various newspapers of the left, centre and right. It can be seen that since 2010 there has been a drastic increase in the presence of racist terms both in left- and right-wing newspapers. This sharp increase may be associated with several factors such as: the causes that led Donald Trump to become president in 2016 or also the exit of Great Britain from the EU, with Brexit. These two events might be linked to the increase in poverty among the middle class, who, feeling less represented, can tend to go in search of the guilty of their condition. It could perhaps be argued that the middle class sees political figures such as Donald Trump or Nigel Farage as the solution to their concerns,

without realizing that these right-wing political movements only divert their attention to new ephemeral problems. It may be no coincidence that immigration and racial discrimination are some of the most frequently discussed topics of recent times and are represented by the establishment as problems, thus increasing hatred and racism towards minorities among the population.

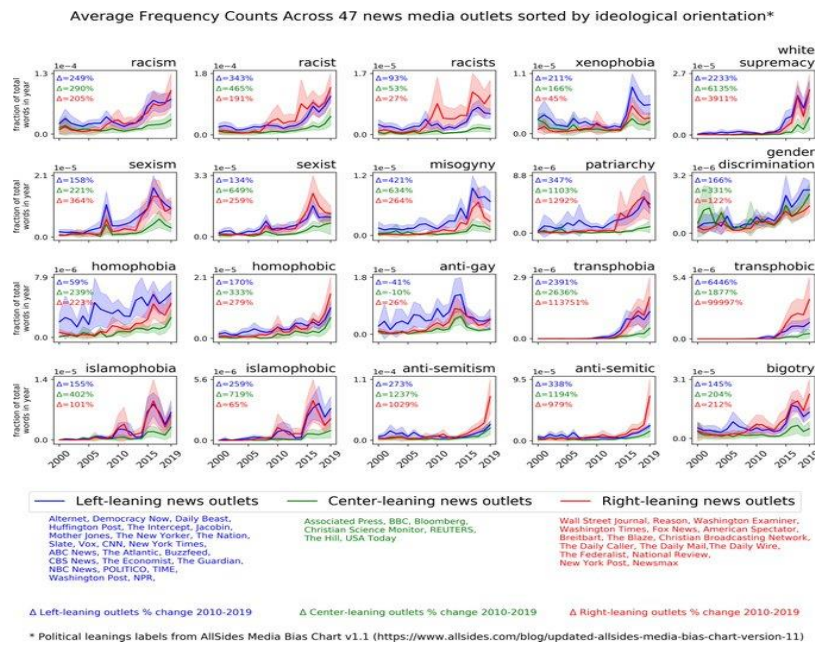


Figure 1: Political leanings labels (The Guardian, 2022)

As regards Busà (2013), provides some examples that suggest how the media can reduce interpretations of what is written to a single possibility. Note in fact how in this title "Villaraigosa calls for high educational standards for black, brown children", an "oversimplified and undiversified" (Busà, 2013: 140) representation of these social groups is given. It is generally argued that the language choices may give the reader the possibility to have only one interpretation of what they are reading. The media might want to shape the interpretation each of us may give on a given subject to their liking. It can be claimed that the media may abuse their power and may influence to discriminate against ethnic minorities, often indirectly. The fact that minorities have less access to discourse sources is only an advantage for the media, which may more easily manipulate them. It is agreed that topics in the media concerning minorities have a negative and

discriminatory tendency, often associated with problems and threats to the white community.

Media attention is also reflected in the visual features of texts and grammatical choices. Regarding the former, "violence and crime of minorities will typically appear in (big) headlines and [...] on the front page" (Van Dijk, n.d: 41). The media, therefore, attach great importance to the visual impact they can exert on the reader. The second aspect, the grammatical one, plays an important role as certain techniques may be used in order to distort the true meaning of the sentence. Some of them are for example: the active sentence (or transitivity) to emphasize the person responsible for the action; nominalizing verbs "such as using the word 'discrimination' instead of saying who discriminated whom" (Van Dijk, n.d: 41); the use of hyperbole or metaphors. The '*swamp*' metaphor (Van Dijk, n.d.) is well known in the UK. It was first used by Margaret Thatcher saying that "British people being 'rather swamped' by people with an alien culture". Examples of hyperbole are: 'huge numbers', 'invasion of immigrants', 'waves of immigrants. In their study Chomsky and Herman (1988) identify five filters that are applied in the reproduction of news and that consequently alter its meaning. Readers, who are used to read filtered news, become increasingly convinced that what is written in the media is 'normalized'. From this statement derives the title of the book as "the public's agreement [...] is not a 'real' agreement, rather, it is a 'manufactured consent'" (Mooney, Evans, 2015: 71).

Among these five filters I would like to focus on three in particular: media ownership, the media elite and the flack. In the first filter, the owner of the newspaper plays a fundamental role. With his/her ideologies and political visions, he/she determines the contents of the media and therefore also selects the target audience. For example, a left-wing paper such as *The Guardian* is expected to have a target audience of mainly left-wing readers². The second filter concerns the power of the establishment over the media.

² How left or right-wing are the UK's newspapers? (<https://yougov.co.uk/topics/politics/articles-reports/2017/03/07/how-left-or-right-wing-are-uks-newspapers>)

They exercise their supremacy by influencing the narrative of the news (Chomsky and Herman, 1988). The third filter is about what happens to those who stray from the consensus, thus becoming inconvenient for the media elite. The term flack represents the elimination and discrediting of such sources (Chomsky and Herman, 1988). In 2020 Goldberg published an article entitled *How the Media Led the Great Racial Awakening* in *Tablet*, a daily online magazine of Jewish news. The graph below shows the percentage use of the terms 'racist' and 'racism' in four major American newspapers

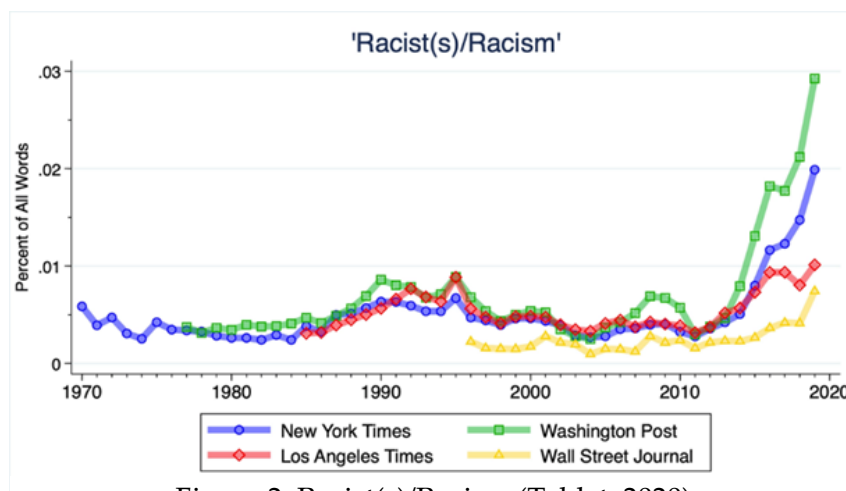


Figure 2: Racist(s)/Racism (Tablet, 2020)

between 1970 and 2019. The last decade has seen an exponential increase in the use of these terms in the press, especially in *The New York Times* and *The Washington Post*. Nowadays, the gap between the former and the latter is enormous. According to FBI reports (FBI, 2021) this big change is due to the increase of racial incidents, hate and racism episodes and the struggles against discrimination carried out by movements such as the Civil Rights movement and Black Lives Matter. One case that caused a huge reaction was the killing of George Floyd in May 2020, following which hundreds of protests broke out across the USA.

1.3 Biased news

As it has been pointed out above, it has been argued that the press may contribute to the spread of racism. One of the ways in which it does so is through the transmission of biased news. According to Van Dijk (2015) this almost always starts with biased newsgathering, i.e. little attention is given to non-white sources, which are considered

unreliable. Secondly, there is the choice of a biased topic. When it comes to minorities, there may be always a tendency to talk about them in a pessimistic way and associate them with negative actions and problems, in order to create negative mental models. The press tends to hide issues that would portray minorities in a positive way (such as minority contributions to the economy, society and culture) and issues such as white racism or discrimination against migrants and asylum seekers (Van Dijk, 2007). "The world of the Press is not the real world, but a skewed and judged world" (Fowler, 1991: 11).

Another aspect not to be underestimated is the use of biased language, i.e. the choice of specific terms and the use of metaphors with negative meaning, hyperboles, racist labels, and biased descriptions (Van Dijk, 2007). In Van Dijk's view, newspaper reports can be said to be "contaminated" by the prevalence given to the white voice and its ideologies. "The selective coverage of 'negative' topics is not only an effect of media bias, but partly also caused by those who feed the press with such stories" (Van Dijk, 2006: 7). We are only shown what the press wants the public to read, a selection process that "gives us a partial view of the world" (Fowler, 1991: 11). Depending on the political orientation of the newspaper, some topics will be given more space than others and each will give a different view. In many of them, however, 'We read more about terrorism of Others against Us, than Our forms of power abuse [...] against Them' (Van Dijk, 2015: 387). The perception of biased news is totally subjective because it depends on the reader's interest in the topic, so depending on that, the news will be remembered more or less easily. However, the reader's exposure to this type of news might lead to the emergence of prejudices and stereotypes against minorities.

1.4 Discrimination against non-white journalists: a connection to Frederick Douglass

As mentioned above, discrimination and racism do not spare ethnic minority journalists either. Although they may be equally qualified as white journalists, the color of their skin or their origin can prevail over their voice and opinion. In some cases, the opinions on issues involving minorities that are most listened to are those of the police and not

those of discriminated journalists (Van Dijk,2006:3). Minorities are considered to be unreliable sources for writing journalistic and online articles, even if they have more knowledge and expertise on a given topic. "Their sources are considered to be less 'objective' and hence less credible" (Van Dijk, n.d.: 215) and they are very rarely allowed to speak on their own, without the control of a 'white' person.

To support this thesis, I will refer to a study carried out by the Reuters Institut in 2019. They sampled 100 newspaper editors (both online and offline) in five racially, historically and culturally diverse markets (Brazil, Germany, South Africa, UK and USA). From the graph, it is clear that editors are overwhelmingly white in relation to the total population, while non-white editors have very little representation. A particular case is South Africa. As it can be seen in the graph below, 68% of the editors are non-white, but it must be taken into account that 91% of the population is non-white. According to the Brazilian population census in 2010, about 50% of the population was non-white, but this study shows that only 5% of editors are non-white. In countries such as Germany and the UK, the percentage of non-white top editors is 0. It is data that certainly gives us food for thought because, even though we are in the year 2022 and that the struggles and demonstrations against racism are increasingly present and frequent, the phenomenon of racial discrimination is still very widespread.

The analysis conducted by Willnat (2017) made a comparison in the US between the number of non-white journalists in the 1970s and the number in the 2000s. The result was quite unexpected: the number of non-white journalists has gradually declined, despite the American population growing in racial diversity.

Figure 1: percentage of non-white top editors by country

Data from 'Race and leadership in the news media 2020: evidence from five markets.'

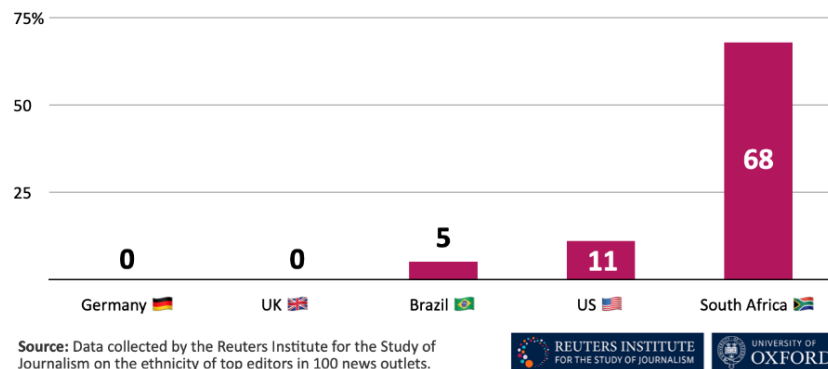


Figure 3: Percentage of non-white top editors by country (2019)

It could be appropriate to mention in this context one of the greatest exponents of the 20th-century slave narrative: Frederick Douglass. In his autobiographical work *Narrative of the Life of Frederick Douglass, an American Slave*, Douglass recounts the years he spent on plantations as a slave. The slaves were kept in a state of complete ignorance and had no access to reading or knowledge. In this work, he recounts the horrors of white wickedness that he had to experience since his childhood. Frederick, therefore, wants to show his humanity and prove that black people can also be reliable narrators. The pen was therefore for him a real instrument of liberation, when neither law nor society offered the same.

In the 20th century, autobiographies of emancipated slaves were seen as a propaganda tool for the abolitionist movement, so their voices were not heard and considered. (Scacchi notes, 2021). As Van Dijk writes in this academic article "Minority representatives will seldom be allowed to speak alone: a white person is necessary to confirm his or her opinion" (Van Dijk, n.d: 39), this can be read as a clear reference to the slave narrative.

1.5 The power of headlines: a case study

Headlines are very important both on the textual level and the cognitive level. "The headline has the important function of grabbing the readers' attention" (Busà, 2013: 80).

They are usually written in bold and must "[...] use condensed, highly synthetic language [...]". (Busà, 2013: 81) in order to best summarise the topic. "They function as explicit instructions on how to construct the overall meaning, macrostructure, or topics of news reports" (Van Dijk, 2015: 388).

From a cognitive point of view, they have an important function: they are the first part that catches the eye of the reader, who will then decide whether or not to read the whole article. The information contained in the headline thus enables the reader to activate a "[...] process of understanding in order to construct the overall meaning" (Van Dijk, 1991: 50). However, headlines highlight the most important things in the journalist's opinion: "They summarise what, according to the journalist, is the most important aspect" (Van Dijk, 1991: 51). It can therefore be said that headlines are a subjective description of what happens in the article. This then leads to influence opinion and reading.

In *Racism and the press* by Teun Van Dijk (1991), an in-depth study was conducted of a corpus consisting entirely of headlines of articles about race, taken from *The Times*, *The Guardian*, *The Daily Telegraph*, *The Daily Mail* and *The Sun*. Van Dijk analyzed the frequency of publication of these types of articles between August 1985 and January 1986 and found that an average of 3 articles per day were published on this subject. Another very important aspect of this study is the use of words and their lexical style "[...] they also signal the social or political opinions of newspapers about the events" (Van Dijk, 1991: 53).

When analyzing the corpus, it was noticed that the first three most common words in the headlines of these newspapers are: police (338 times) and its synonyms, riot (320 times) and black (244 times). It can be seen that instead of using less negative words such as disturbances or unrest, the press prefers to use the term *riot*, to emphasize this act as a violent and criminal act. "Words such as 'ethnic groups' or 'minorities' hardly occur [...] for the British Press are primarily seen in terms of racial categories, that is in terms of color or appearance" (Van Dijk, 1991: 55). This would appear to suggest, therefore, that

the British press favors the use of racist terms to identify minorities. Headlines of the coverage of 'race riots' focus mainly on black violence than on police violence. White violence is occasionally associated with negative terms.

The use of active and passive agents (transitivity) for the transmission of meaning is very important. Almost always minorities are active agents committing acts of violence, whereas the police are often the passive agents suffering the action. The study found that 'police are agents of a negative action in only 7 out of 171 cases', whereas 'minorities are agents of negative actions twice as often' (Van Dijk, 1991: 61). From this study it can therefore be deduced that, especially in the right-wing press, violence committed by blacks features heavily in the headlines, compared to white violence. As Van Dijk (1991) writes, the police are often depicted as agents of neutral or positive actions, while blacks are almost always responsible for negative actions.

Conclusion

This chapter has presented and discussed several domains, in which extensive research on the concept of racism and discrimination has been carried out. I first explored how discourses and the media may play a very important role in the dissemination of this phenomenon, which often occurs indirectly and 'hidden' from the reader's eye. I gave examples how linguistic choices can influence the intrinsic meaning of a sentence and thus lead to the birth of the so-called 'biased' news: news reworked through specific techniques and the creation of negative mental models in the minds of readers. Chapter 1 dedicated a paragraph to a current topic, which still recalls the colonial past and slave narrative, that is discrimination against non-white journalists. The last part analyzed a case study included in the book *Racism and the press* by Van Dijk, which highlights the importance that headlines can have in the cognitive and textual perception of the text. The analysis of the corpus showed that the most recurrent words were all related to negative aspects of racial discrimination.

2. Corpus linguistics in discourse analysis

This second chapter is divided into four sub-sections with the primary purpose of describing the importance of corpus analysis in the study of a specific language. I will start by introducing what corpus linguistics is and its main characteristics, focusing on the concept of qualitative and quantitative analysis. I will then talk about how to build a corpus, in which it is essential to make a selection of texts that are coherent with the topic under analysis. Another aspect I will touch on is the choice of corpus size and annotations which are a widely discussed topic. As the chapter progresses it will come closer and closer to what will be covered in the third and final chapter. In fact, in the third section, I will deal with three fundamental aspects of the analysis of a corpus namely: frequency, concordance and collocation; and then end with the fourth section in which I introduce the use of a reference corpus citing a case study (Weir, Anagnostou, 2007).

2.1 Corpus linguistics

To introduce this topic it is necessary to provide, first of all, a definition of *corpus*. “Corpus is a term from the field of linguistics and refers to a large set of texts [...] which is considered to be representative of a language and is used to analyze it.”³ Corpora may be of various types such as multilingual or monolingual, they may contain texts from different historical periods or from the same period (synchronic or diachronic), or they may contain texts from the same field, such as medical texts, etc. (See footnote 1) “Corpora are great resources for observing language as it is used in real communicative situations” (See footnote 1). Corpora are therefore very important tools for analyzing in-depth various aspects of a particular or specialized language in a certain field. McEnery & Wilson (1996:1) provide a clear definition of corpus linguistics: “the study of language based on examples of real life language usage”. “Corpus linguistics uses large collections of both spoken and written natural texts that are stored on computers [...] By using a

³ FutureLearn, *What is a corpus?* (<https://www.futurelearn.com/info/courses/understanding-dictionaries/0/steps/62809>)

variety of computer-based tools, corpus linguists can explore different questions about language use.” (Schmitt and Rogers, 2020: 91). Corpus linguistics is really useful to explore patterns of language use and “can provide tremendous insights as to how language use varies in different situations, such as spoken versus written, or formal interactions versus casual conversation”. (Schmitt and Rogers, 2020: 91) Overall, the most important contribution of corpus linguistics is that it can bring together aspects of quantitative and qualitative techniques (Schmitt and Rodgers, 2020).

Biber et al. (1998) argue that this type of study has four main characteristics including: an empirical approach to analyzing the use of language patterns, a “representative sample of the target language” as the basis of the analysis, extensive use of specific software to identify language patterns and, lastly, CL relies on both qualitative and quantitative analysis to interpret the results.

A new discipline that has become increasingly popular in recent years is Corpus-Assisted Discourse Analysis (CADS). The purpose of this discipline is described by Partington (2008: 96) as “the investigation and comparison of features of particular discourse types, integrating into the analysis where appropriate techniques and tools developed within corpus linguistics”. To carry out an analysis using CADS approach, it is necessary “to acquaint ourselves as much as possible with the discourse types in hand” (Partington, 2008: 98).

I will now explore the difference between qualitative and quantitative analysis. The first analysis focuses more on the use of the language in detail. It does not pay attention to frequencies, numbers and data. This analysis makes it possible to identify ambiguities in the language, such as detecting double meanings of a certain word. An example is the word "red", which can mean either the color red or a political aspect such as communism. Thus, in the word "the red flag", the qualitative analysis attributes both meanings. The disadvantage of this analysis is that, since it is more subjective, the results are not analyzed statistically so it is not known whether they are actually significant. The second analysis, the quantitative one, is more objective, so its results are statistically reliable. It

is based more on numbers, data, and frequencies. With this analysis, it is possible to identify the linguistic phenomena that are part of a language and those that are simply random events. However, the results of such an analysis are less inclusive, since, as in the example "the red flag", one has to choose whether to classify "red" as "political" or "color".⁴

2.1.1 Corpus building

The building of a corpus is a very important step in the process of linguistic analysis. Baker (2006: 25) writes: "The process of finding and selecting texts, obtaining permissions, transferring to electronic format, checking and annotating files will result in the researcher gaining a much better 'feel' for the data and its idiosyncrasies". To start building a corpus, it is necessary to have a research question and this will serve as the basis for the building of the corpus. Furthermore, the corpus must be representative of the type of language being analyzed, e.g. language in newspapers, spoken conversations, academic language (Anne O'Keeffe, Michael McCarthy, 2010). Another fundamental aspect in building a corpus is *sampling*. "Many corpora are composed of a variety of texts, of which samples are taken. [...] This technique of sampling is in place to ensure that the corpus is not skewed by the presence of a few very large single texts taken from the same source" (Baker, 2006: 27).

When creating a corpus, it is essential to select texts that are suitable and consistent with the topic to be analyzed. If no attention is paid to this step, results may be considered not reliable. Baker (2006: 26-30) identifies three different types of corpus: specialized corpus, diachronic corpus and reference corpus. A specialized corpus "would be used in order to study aspects of a particular variety or genre of language" (Baker, 2006: 26). Such a corpus, therefore, focuses on a single type of writing, from academic to journalistic. In this case, it is not necessary to create a corpus consisting of millions of words, because "we may want to be more selective in choosing our texts, meaning that

⁴ *Qualitative vs Quantitative analysis* (<http://www2.sal.tohoku.ac.jp/ling/corpus3/3qual.htm>)

the quality or content of the data takes equal or more precedence over issues of quantity.” (Baker, 2006: 29). A diachronic corpus “is a corpus which has been built in order to be representative of a language or language variety over a particular period of time, making it possible for researchers to track linguistic changes within it” (Baker, 2006: 29). The use of this type of corpus is very useful when one wants to analyze the linguistic changes in a given linguistic field, which occurred over a certain period of time. The third type is reference corpus which “consists of a large corpus [...] which is representative of a particular language variety” (Baker, 2006: 30). An example cited by Baker is the *British National Corpus* (BNC). It is a corpus composed of one hundred million words of written and spoken data. Like this one, there are many others that are publicly accessible and thus allow linguistic analyses to be carried out with reference to very large and inclusive corpora.

2.1.2 Corpus size and annotation

Size and annotation have always been a much-discussed topic with often conflicting opinions. “The optimal size of a corpus can be reached only when the collection of more texts does not shed any more light on its lexicogrammatical or discourse patterning” (Baker et al., 2008: 275). The size of the corpus is a very important element in conducting a linguistic analysis. In the last fifty years, corpus size has increased considerably because it was noticed that in smaller corpora it was more difficult or sometimes impossible to detect certain linguistic features. (Anthony, 2013: 145). If a corpus is small, it will only be possible to analyze a few examples of the linguistic phenomenon in question. Sinclair (2004: 189) argues “If within the dimensions of a small corpus [...] you can get the results that you wish to get, then your methodology is above reproach – but the results will be extremely limited”. Essentially, what makes a corpus valuable and useful is its content and the results it can offer us, not its size.

Corpus annotation is an issue that has both supporters and non-supporters. Those who are against argue that “it contaminates the original data making it more difficult to see language patterns” (Anthony, 2013: 147). They believe that annotation is an element that

overlaps with pre-existing language patterns, thus disturbing the search for new language patterns in the corpus. Instead, those who support this technique consider “the addition of annotation as a necessary step in order to test a particular linguistic theory.” (Anthony, 2013: 148). Furthermore, with the latest technologies, it is possible to hide annotations when analyzing a corpus and they can also provide additional information in the search for linguistic features.

2.2 Corpus-based and corpus-driven approaches

Studies with corpora have two schools of thought: corpus-based studies (the study based on corpora) and corpus-driven studies (the study induced by corpora). The first one is described by Tognini-Bonelli as follows: “The approach corpus based is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study” (Tognini-Bonelli, 2001: 65). In corpus based studies “it is impossible to completely remove all pre-existing ideas about language before observing corpora” (Anthony, 2013: 142). It is very important to distinguish between the top-down and bottom-up approaches. The first approach starts with the general structure and breaks it down to its elementary components. The second approach starts with the basic elements and goes on to create a more complex system.⁵ The corpus driven approach, instead, considers the corpus data in their entirety in order to draw conclusions⁶. In this approach “direct observations of the corpus should be the starting point of analyses” (Anthony, 2013: 142).

Biber et al. (1999) conducted a study based on a corpus of 40 million words using the corpus-driven approach. With this approach it is, therefore, easier to identify the structure of the language and above all its lexical patterns, which are identified by analyzing their frequency in a corpus. According to Biber et al. (2009: 81) the corpus-

⁵ *Differenza tra bottom up e top down?* (<https://qualcherisposta.it/differenza-tra-bottom-up-e-top-down>)

⁶ *Gli approcci corpus based e corpus driven* (<https://123dok.org/article/gli-approcci-corpus-based-e-corpus-driven.dzxejwvq>)

driven study needs to have three fundamental characteristics: it must analyze the sequences of lexical forms, it must focus on the most frequent combinations of lexical forms and must be based on the analysis of the true lexical forms in the corpus. (See footnote 3 above)

2.3 Frequency, Concordances and Collocates

When analyzing a corpus there are three main aspects that need to be investigated: frequency, concordance and collocation. “Frequency is one of the most central concepts underpinning the analysis of corpora” (Baker, 2006: 47). A good starting point to analyze a corpus is a frequency list, which “can help to give the user a sociological profile of a given word or phrase enabling greater understanding of its use in particular contexts” (Baker, 2006: 47). Analysis software makes it possible to create a *word list* which is “a list of all of the words in a corpus along with their frequencies and the percentage contribution that each word makes towards the corpus” (Baker, 2006: 51). With a word list is possible to see also the ratio between types and tokens. If the ratio is low the corpus will have big repetitions, if the ratio is high the language of the corpus will be more diversified. The type/token ratio is important because it can help the user to see how complex and specified the corpus language is.

Most of the time the most repeated words in a corpus are function words such as ‘the’, ‘and’, ‘to’, ‘of’, ‘for’. “Such words belong to a closed grammatical class [...] these categories tend not to be subject to linguistic innovation” (Baker, 2006: 53). One aspect to consider when making word lists is ‘*lemmas*’, which are “canonical forms of a word” (Baker, 2006: 55). While analyzing a small corpus, Baker (2006) noticed that in the list of the 10 most frequent lexical words, the words ‘bar’ and ‘bars’ both appear. That is why it is important to recalculate word frequencies including lemmas. Within the topic of frequency, clusters should not be forgotten. In corpus analysis software, it is possible to specify the size of the cluster in order to see which words occur most often in the cluster containing a specific term. “Examining frequent clusters of words or their dispersions

across a text may be more revealing than just looking at words in isolation" (Baker, 2006: 68).

We can continue with concordance which is "simply a list of all the occurrences of a particular search term in a corpus, presented within the context that they occur in, usually a few words to the left and right of the search term" (Baker, 2006: 71). Baker (2006) carries out a concordance analysis which is "one of the more qualitative forms of analysis associated with corpus linguistics" (Baker, 2006: 89). He uses *WordSmith Tools* and relies on a newspaper articles corpus which includes the words *refugee* or *refugees*. "The object of creating concordances is to look for patterns of language use, based on repetitions" (Baker, 2006: 77). The first step for the analysis is to focus on a specific keyword as, in this case, *refugee*. Once selected the first word to be examined, the research can start. The analysis tools will provide concordance lines and each of them will "show an occurrence of the search term, with a few words of context either side of it" (Baker, 2006: 77). It is very useful to sort concordances in various ways to the left or to the right of the key word. The first result obtained by Baker (2006) is that in this corpus refugees were described mainly in terms of huge and problematic quantities, thus in a negative light. Furthermore, the key word was often followed by verbs such as 'to flee', 'to return', 'to deliver', 'to overflow'. "This pattern suggests that refugees are often described in terms of their movement" (Baker, 2006: 80).

Other relevant aspects that came up in the analysis were many patterns that described refugees "in a way that connected them to tragic circumstances" (Baker, 2006: 83).

One of the last steps of concordance analysis is "Attempt to hypothesize why the patterns appear and relate this to issues of text production and reception" (Baker, 2006: 93).

The last aspect is collocation. "When a word regularly appears near another word and the relationship is statistically significant in some way, then such co-occurrences are referred to as collocates and the phenomena of certain words frequently occurring next to or near each other is collocation" (Baker, 2006: 96). Collocation is used to gain an in-

depth understanding of the position of certain words under investigation, starting from the context in which they are found. The easiest way to calculate collocations is done by counting the number of times that a specific word appears in the selected span (for example three words to the left and three words to the right) of the search word.

However, some problems may be encountered, as by using this method the focus is not on the collocation with respect to a particular term, but only on the calculation of the frequency of the term, which can be located near words other than the term under investigation. A test called MI (Mutual Information) was therefore created which “expresses the extent to which words co-occur compared to the number of times they appear separately. MI Score is affected strongly by the frequency, low-frequency words tend to reach a high MI score which may be misleading.”⁷

A collocational analysis is useful because “it provides a focus for our initial analysis which is particularly helpful when a large number of concordance lines needs to be sorted [...] in order to reveal lexical patterns” (Baker, 2006: 114). Furthermore, collocation helps us to see “the most salient and obvious lexical patterns surrounding a subject.” (Baker, 2006: 114). It is also beneficial to examine “the historical backgrounds and etymologies of words” (Baker, 2006: 119), in order to have a wider vision of the meanings a word can have.

2.4 Using a reference corpus

In this last section, I want to illustrate the benefits of using a reference corpus in conducting language studies and analyses. As a basis for a language study, corpora are a great advantage because firstly “are representative enough of a particular genre of language” (Baker, 2006: 43) and thus make it possible to select the field of research;

⁷*Sketch Engine* (https://www.sketchengine.eu/my_keywords/mi-score/#::~text=The%20Mutual%20Information%20score%20expresses,score%20which%20may%20be%20misleading)

secondly, they make it possible to analyze in-depth the various linguistic aspects and phenomena and, above all, it “acts as a good benchmark of what is ‘normal’ in language” (Baker, 2006: 43). For a corpus to be effective, it must be large enough to represent most of the relevant varieties of the language. It is therefore important to try to include as many language varieties as possible in order to obtain reliable results. Therefore, corpora serve to show the most recurrent linguistic phenomena and how language structures are used in a given field. In order to support these statements with real data I would like to report a case study (Weir, Anagnostou, 2007), which analyses and compares two corpora: the first consisting of about 30,000 articles published in 2005 in a Scottish daily newspaper and the second the BNC (British National Corpus), a reference corpus which

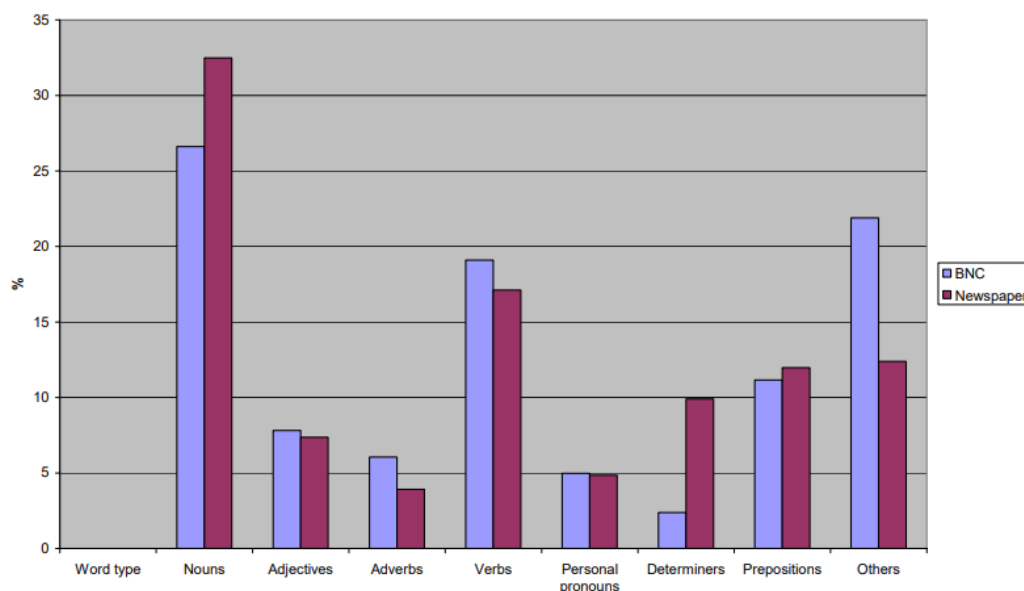


Figure 4: Comparison of Newspaper Corpus and BNC by % POS

“covers British English of the late 20th century from a wide variety of genres, with the intention that it be a representative sample of spoken and written British English” (Wikipedia - the free encyclopedia).

In the graph above the comparison between the various POS (Part of Speech) of the two corpora is represented. In particular, there are two categories that show unexpected results. The first one is the 'nouns' column which shows that in the newspaper corpus the % of nouns is about 6% higher than in the BNC. The second is the 'determiners' column, which in the newspaper corpus are about five times the number of the BNC.

However, upon further analysis, the researchers noticed that in the newspaper corpus there were numerous typing mistakes by the POS tagging software and also in the source data (e.g. missing spaces between words). With this study it was, therefore, possible to identify some errors within the newspaper corpus, which in fact altered the results (see Figure 4); but also to make comparisons with a large reference corpus.

In the third and final chapter of this dissertation I will carry out a linguistic analysis using a corpus I have created. Having created this corpus consisting entirely of articles relating to the '*Black Lives Matter*' activist movement, it will allow me to actually be able to see how the influence of the 'white' press is exerted in the representation of ethnic minority movements. In addition, it will be possible to see the differences between the British and American press in the use of language terms and structures.

Conclusion

This chapter has dealt with several issues concerning corpus linguistics in discourse analysis and the study of language. I first explored the definition of corpus linguistics, focusing on the difference between qualitative and quantitative analysis. The chapter went on to explore more practical aspects such as: corpus construction, which is the fundamental part of a good linguistic analysis; the size of a corpus and its importance and annotations. This chapter is intended as a preface to the third part. In fact, in the third and fourth sections I have illustrated two topics that anticipate what will be discussed in the final part of this thesis, also giving a concrete example of a case study proposed by George Weir and Nikolaos Anagnostou. In this chapter, therefore, I have provided the tools and concepts necessary to carry out, in the next chapter, the analysis of the corpus I have created.

3. Corpus linguistics analysis of Black Lives Matter

In this last chapter, I will carry out a corpus linguistics analysis by using a corpus I created by myself. This corpus is composed of articles related to the topic of racial discrimination and racism, in particular articles that deal with the activist movement for the rights of Afro-Americans: Black Lives Matter. It is a monolingual and diachronic corpus, because the articles I collected were published in different years. In this analysis, I will use the corpus-driven approach, which considers the corpus data in their entirety in order to draw conclusions (Anthony, 2013: 142). This approach will make it easier to analyze the lexical patterns, the language structures and the different lexical forms.

In compiling the corpus, I wanted to collect articles from both the British and American press in order to be able to examine the differences between the two countries' views on this topic and especially to analyze the use of language terms and structures. In the table below you can see the data concerning the general corpus containing all articles.

Number of articles	51
Sources	The Telegraph, The Guardian, The Washington Post, USA Today, The Los Angeles Times, Metro UK, New York Post
Dates	From 2020 to 2022
Word Tokens	49,718
Word Types	6719

Table 1: General corpus data

I then created sub-corpora, one for each newspaper, which would be useful when analyzing the differences between them. In the table below, you can see the data of the individual corpora. To conduct this linguistic analysis, I will use the free software AntConc, designed specifically for use in the classroom.

	The Telegraph	The Guardian	The Washington Post	USA Today	The Los Angeles Time	Metro UK	New York Post
Number of articles	2	10	10	4	8	10	7
Dates	2020-2022	2020-2022	2020-2022	2020-2022	2020-2022	2020-2022	2020-2022
Number of words	1,397	11,890	13,954	3,444	6,923	6,073	5,081

Table 2: Sub-corpora data

3.1 Word frequency

The first element that, if analyzed, can give us some clues is certainly word frequency (Baker, 2006). The analysis of word frequency begins with the creation of the word list, which includes both lexical and function words. As I have already mentioned in Chapter two, the word list is “a list of all of the words in a corpus along with their frequencies and the percentage contribution that each word makes towards the corpus” (Baker, 2006: 51). I will also take into account the cluster sizes (most of the time three words to the left and three words to the right) in order to see which words occur most often in the cluster containing a specific term.

My analysis, therefore, starts with the general corpus. From the frequency list, I chose the first fifteen most frequent lexical words and analyzed their normalized frequency per 10,000 words, as it is shown in the table below.

Word	Frequency	Normalized Freq. per 10.000 words
Black	473	95
Police	415	83
Lives	273	54
Matter	250	50

People	222	44
Officers	142	28
City	140	28
Protesters	125	25
Protests	117	23
Floyd	96	19
Protest	93	18
Officer	85	17
BLM	83	16
Movement	77	15
White	75	15

Table 3: The first 15 most frequent lexical words in the general corpus

The result is not surprising: the first two most recurrent words are in fact 'black', followed immediately by 'police'. Furthermore, in this list, it can be seen that no less than three of the fifteen words refer to the police, as there is the synonym 'officer' and its plural 'officers'. This shows us how articles that are supposed to deal with the topic of the fight for the rights of black people always tend to include the figure of the policeman and in such a way as to highlight their struggle as a danger and a threat that therefore requires the protection of the police (Van Dijk, 1991). Another result I would like to focus on is the word 'riot', which occurs 25 times in the general corpus. It is very common for newspapers to prefer to use terms such as 'riot', which have a mainly violent and negative connotation, even to describe peaceful demonstrations instead of using less negative words such as 'demonstration' or 'unrest', as I already mentioned in Chapter One (Van Dijk, 1991).

My analysis now continues with some sub-corpora. We can start with *The Guardian* newspaper. From the frequency list, I reported the fifteen most frequent lexical words and their normalized frequency per 10,000 words.

Word	Frequency	Normalized Freq. per 10,000 words
Black	114	95
Police	114	95
Lives	65	54
People	61	51
Matter	57	47
Protests	37	31
City	36	30
Brown	33	27
Movement	33	27
Group	32	26
Protest	31	26
Officers	30	25
Ferguson	29	24
Protesters	27	22
March	26	21

Table 4: The first 15 most frequent lexical words in *The Guardian* sub-corpus

Here too, we note that the first two most frequent words are always 'black' and 'police', among other things with the same number of occurrences. So, we can deduce that there is also a tendency in the British press to always mention the issue of justice. In this list two words in particular attracted my attention: 'Brown' and 'Ferguson'. On viewing the concordances of both words, I discovered that the word 'Brown' is almost always used as a surname and not as a color. In fact, both 'Brown' and 'Ferguson' refer to the same incident of violence against black people. They both refer to the killing of the black teenager Michael Brown that took place in August 2014 in Ferguson, Missouri. Reading the articles about this I noticed that this episode of violence was the drop that broke the camel's back, as it is written in this article in *The Guardian*: "The killing of Michael Brown

created a new generation of black activists, with thousands taking to the streets, and a hashtag used more than 27m times" (Lowery, 2017).

The Washington Post sub-corpus, instead, shows slightly different results. The fifteen most recurrent lexical words are reported in the table below.

Word	Frequency	Normalized Freq. per 10,000 words
Police	126	90
Black	112	80
Lives	71	50
Matter	71	50
People	61	43
City	45	32
Year	44	31
Officers	37	26
Protests	35	25
Craighead	33	23
Protesters	32	22
Public	30	21
Protest	27	19
Brooks	25	17
Floyd	25	17

Table 5: The first 15 most frequent lexical words in *The Washington Post* sub-corpus

The Washington Post is classified as a center-left newspaper, hence mainly liberal. Here too, as in *The Guardian*, the first two words are 'black' and 'police', but it can be seen from this list that 'police' is used more times than 'black', which is supposed to be the main subject, whereas in *The Guardian* it is the other way around: 'black' is more frequent.

There is also another concept to be noted: minorities are often described by their skin color (e.g. 'black people', 'black protesters', 'black migrants'). This makes us realize that skin color is the most commonly used feature to describe an ethnic minority. Instead of using terms such as 'ethnic groups' or 'minorities', there is always a tendency to focus on skin color, as if all the characteristics of an ethnic minority could be included in a simple qualifying adjective. In this list, we can notice two personal names: Floyd and Brooks, both with the same frequency (25). Reading the articles in the corpus I found out that 'Floyd' refers to the police killing of the black man George Floyd that happened in 2020 in Minneapolis, which I will cover later; while 'Brooks' refers to a black man, which "was arrested in June on charges of inciting a riot and resisting a police officer during a Black Lives Matter protest." (Craig, 2021). This BLM protest in particular was made in response to the killing of Floyd. From this, we can see how despite the killing of a black man, violence and discrimination continue to occur like a chain reaction: one protests to give justice to a murdered comrade, and then one is unjustly arrested.

3.2 Collocations and concordances

The second feature I will focus on is the collocation of specific search terms. The AntConc Collocates tool allows you to view words that often appear in close connection or close association with the search term. A collocation analysis is very helpful because it can show us how minorities are described and with what terms they are most often associated. (Baker, 2006). As I have already mentioned in Chapter 2 this kind of analysis can help to identify "the most salient and obvious lexical patterns surrounding a subject" (Baker, 2006: 114).

I will focus on search terms that are closely related to the topic of racial discrimination and the fight for freedom, first using the general corpus and then using sub-corpora of both British and American newspapers. I will use both the three words to the right and three words to the left span, which are the most commonly used and I will sort them by frequency.

The analysis starts with the general corpus. The first search term is 'black'. The lexical word that is placed several times next to 'black' is 'police'. If we go deeper into the analysis, we will notice, as in the image below, that three out of seven sentences deal with the topic of police violence against the black population.

ng. Handmade posters advocated "Fund services not police" and "Amplify black voices." Another read: "Black stayed the same. Since Floyd's murder, police have killed Black and brown men in Side — at least at first. In Columbus, police killed a Black teenage girl. That fits negotiators was a targeted retribution for the police killings of black men. A week later, that read 'Decolonise everything' and 'Defund the police'. Protestors shouted 'Black lives matter every day' that read 'Decolonise everything' and 'Defund the police'. Protestors shouted 'Black lives matter every day'

Figure 5: Concordance of the term 'police'

I then tried changing the search term and searched for the word 'unrest', which would be a term with a less negative connotation to describe a peaceful protest and it occurred only 9 times. If instead one searches for the term 'riot', which has a more negative connotation, then more results come up: as I have already said, 'riot' occurs 25 times in the general corpus, which is more than the double. Therefore, it can be deduced that the media generally tend not to report protests as peaceful, but to report them with a negative connotation. I also noticed that there are several references to the 'Capitol riot', which is the attack on Capitol Hill carried out by Trump supporters on the 6th of January 2021 to overturn his defeat in the US presidential elections. The word 'Capitol' occurs 31 times in the corpus, which may suggest that this event had quite extensive media coverage.

The next search term is 'power' because I want to see if in some articles the topic of power abuse is written about. In the collocates of 'power', with two words to the left and two to the right span, we can find the word 'police' and 'abused'.

bills that would give police more power to arrest and detain protesters. Florida's new "anti-rioting" law with the department abused their power by indiscriminately arresting people just for being outside,

Figure 6: Concordances of the term 'power'

These sentences show the context in which the word 'power' is placed, but with different meanings. The first sentence speaks of a bill that intends to give even more power to the police in situations of demonstrations and thus take away even more of the freedom to protest. The second sentence speaks of the real abuse of power by the Santa Monica police department. In this sentence, the police used their power illegally, arresting people who were participating in a peaceful protest over the death of George Floyd. Once again, this chain reaction (which I have already mentioned in the previous section) occurs, whereby the African-American population is the victim of injustice in participating in peaceful protests.

Another term I would like to focus on is 'knee'. When one hears this word, the incident of police violence in Minneapolis, which resulted in the death of George Floyd, certainly comes to mind. The death was caused by a white policeman who stopped Floyd's breathing with his knee. Listed below are the words that are placed next to the word 'knee': *take, taking, took, pressing, officer, police*. In the figure below, we can see two different uses of the expression 'taking the knee'. In Figure 7, we notice how the noun 'knee' is always prefixed by the verb 'to press', and followed by 'into his/George Floyd's neck'. It is therefore evident that in this case the expression is used to indicate the horrible act of violence committed by the police officer.

Minneapolis police officer Derek Chauvin pressing his knee into George Floyd's neck as Floyd cried for his mother th
 died after a white officer held him down by pressing a knee into his neck for several minutes on May 25, sparking da
 athe' after a white officer held him down by pressing a knee into his neck (Picture: Getty) The police leaders' stateme
 Minneapolis police officer Derek Chauvin pressing his knee into George Floyd's neck as Floyd cried for his mother
 died after a white officer held him down by pressing a knee into his neck for several minutes on May 25, sparking c

Figure 7: Concordances of the term 'knee'

If we look at Figure 8, however, we can see that the term 'knee' is used with a different meaning. The expression 'taking the knee' is "a symbolic gesture against racism whereby an individual kneels upon one knee in place of standing to attention for an anthem or other such occasion" (Wikipedia). In fact, the sentences I have reported speak precisely of symbolic gestures made during anti-racist protests. One can see how in the third sentence the action is carried out by the 'London police officers', who thus show respect

and support for the BLM movement, towards these horrific acts committed by their 'colleagues'.

Demonstrators take part in the Take A Knee for George Floyd solidarity protest organised by Stand Up To Racism Scotland, 'Taking the knee is a personal decision and should individual officers at their own discretion, London police officers were captured taking the knee alongside protesters at a Black Lives Matter demonstration

Figure 8: Concordances of the term 'knee'

The analysis now continues with the sub-corpora. For the British press, I will compare the sub-corpora of *Metro UK* and *The Guardian*. I will continue using the three words to the right and three words to the left span and sort them by frequency. I will begin with *Metro UK*, which is considered a right-wing newspaper. The first search word is 'black'. The most frequent collocates of 'black' are: 'movement', 'protest', 'protesters', and 'minority'. I would like to focus on the term 'protest'. As it is shown in the figure below the word 'protest' is always directly connected to the Black Lives Matter movement. This newspaper describes the struggles for the rights of Afro-Americans as protests, which certainly has a negative connotation.

uters) People take a knee during a Black Lives Matter protest rally at Windrush Square, Brixton (Thousands gather at Hyde Park for Black Lives Matter protest <https://metro.co.uk/2020/06/03/thousands-6/03/thousands-gather-hyde-park-black-lives-matter-protest-12798955/> Thousands of people are protesting in London and in the city.' Five arrested in Black Lives Matter protest over George Floyd's death <https://metro.co.uk/2020/05/31/five-arrested-black-lives-matter-protest-outside-us-embassy-12784215/> Thousands h

Figure 9: Concordance of the term 'black'

I will now change the search term to the word 'kill*'. Using the asterisk at the end of the word, AntConc gives all words or verbs starting with 'kill-' as results. Words that begin with kill- are 9 in this corpus. From the results, we can see that 4 out of 9 are related to the killing of Floyd, which was widely reported in the media all over the world. Despite this, we can see that all other results are still linked to violent actions against black minorities.

was steeped in his own racist world views. 'That man stopped and killed a 17-year-old boy. And then he was acquitted of the murder. to be quite honest, about the killing of Trayvon Martin. Trayvon was killed by a man who was a self-appointed security guard, but really Lives Matter began seven years ago after an unarmed Black child was killed for no reason at all. 'What I was feeling at the time leads to shooting, and that's why a man was shot and killed in Minneapolis on Wednesday night – or look at what just happened in wait for anyone,' says Opal. 'You don't get to people being killed in the street with impunity overnight. 'But also, that means things are been at the forefront of our minds in the wake of the killing of George Floyd. Protests against systemic racism and police brutality hav communities and culture committee, said: 'In the wake of the sickening killing of George Floyd (in the US) the global calls for change and communities and culture committee, said: 'In the wake of the sickening killing of George Floyd (in the US) the global calls for change and of pain,' says Opal, 'and rage to be quite honest, about the killing of Trayvon Martin. Trayvon was killed by a man who was a

Figure 10: Concordances of the term 'kill-'

I will now focus on *The Guardian* sub-corpus, which is considered a center-left newspaper. The first search term here will be 'officer'. In this corpus, we get 23 results for this word. As it is shown in the figure below the word 'officer' is located in the same context most of the time.

evidence to charge Ferguson police officer Darren Wilson with a crime in the killing of Mike Brown, I soothe his autistic patient, when an officer fired his gun three times. "I was thinking as long as I have my the following summer, when a police officer named Darren Wilson shot and killed 18-year-old Michael Brown render or attempting to attack Wilson when the officer shot him. They did know that the police in Ferguson Cleveland police had hired Loehmann, the officer who shot Rice, without checking his references or running

Figure 11: Concordances of the term 'officer'

I believe that if the same linguistic analysis were carried out on another human rights movement (perhaps not of an ethnic minority), I am almost certain that the results would not be so similar and repetitive. It is very significant that in every article about the BLM movement there is at least one reference to police violence because this shows that despite their fights against racism and for equal rights, there is no end to violence and injustice.

I will continue by searching for the word 'kill*'. Words that begin with kill- are 28 in this corpus, so here we get more results compared to *Metro UK* newspaper. In the figure below I report some of the results and it can be noticed that almost all the outcomes refer to killings of black people, in fact, you can even notice two names of people, which appeared several times in my research (George Floyd and Michael Brown).

s a targeted retribution for the police killings of black men. A week later, another lone wolf attacked officers in nt was set up in the US following the killing of black teenager Trayvon Martin in Florida four years ago. NYPD rally in Redwood City after the police killing of George Floyd. The city recently had a Black Lives Matter mural : es-matter-birth-of-a-movement The killing of Michael Brown created a new generation of black activists, with . Covering Ferguson directly after the killing of Mike Brown involved hours on the streets, with clusters of repo :er Darren Wilson with a crime in the killing of Mike Brown, I had been in that city for the better part of three n so much at stake." While the targeted killings of the officers in Dallas and Baton Rouge prompted some commi ng on race that stretched far past the killing of unarmed residents – from daily policing to Confederate imager

Figure 12: Concordances of the term 'kill-'

We can move on with concordance. The AntConc Concordances tool allows you to analyze the context in which specific words are collocated and shows you the kind of patterns that the search word appears in (Baker, 2006). The concordance analysis is beneficial because it helps to understand in which context the search terms are located most of the time. I will begin this analysis with the general corpus by searching for the three words to the right of the keyword. 'Black' is the main search term we are comparing, so we will start with this. In the general corpus, we can find 473 hits of the word 'black'. I would say that 90% of them are composed of 'black' followed by 'lives matter', but this is not surprising since all articles talk about this movement.

However, there are some concordances that show the word 'black' followed by 'and brown', so the skin color is specified here. If one reads the entire sentence, one can see that all five results speak of gun violence and killings of the brown and black population by the police.

ers whose actions were at worst criminal and at best lacked racial sensitivity, and **black and brown bodies** disproportionately gunned down by those sworn to serve and protect; how will police respond? While Chauvin was convicted of murder, the killing of **Black and Brown citizens** by police has continued, often captured on video. In recent weeks, it is hopeful that more New Yorkers will stand up and show their outrage when innocent **Black and Brown individuals**, many of whom are children, are victims to gun violence," he said. But too much has stayed the same. Since Floyd's murder, police have killed **Black and brown men** in Georgia and California. Around the country, six Black people have been killed. A bill passed that will suddenly encompass all of the ways the system marginalises **black and brown people**. We have to redo the whole damn thing." Many of the young activists

Figure 13: Concordances of the term 'black'

The word 'riot' appears 25 times in the general corpus. As we can see in the figure below 1/5 of the hits talks about 'riot gear', so it may suggest that police officers may fear the protests of this movement, which most of the time, are pacifist protests.

days, police began behaving similarly no matter where the protests were. I saw fewer officers in **riot gear and more** on bikes. My theory is that local elected officials or the Chicago police made mistakes. The Ferguson and St Louis County police had sent scores of officers, some in full **riot gear and tactical** vehicles, to deal with the growing crowds and to hold them back as they made an impromptu march to the other side of town. It wasn't much later that the **riot-gear-clad officers** entered the McDonald's, suggesting we all leave because, with protests still simmering in downtown L.A., in the Fairfax district, in Santa Monica and other areas as police in **riot gear formed skirmish** lines opposite protesters and others in the crowds burglarized stores and set them on fire. Looting, police cars and government buildings set on fire and monuments vandalised. Officers in **riot gear have been** blasting activists with rubber bullets and tear gas, and were even filmed using

Figure 14: Concordances of the term 'riot'

In this analysis, for the American press, I will compare the sub-corpora of the *Los Angeles Times* and *The New York Post* newspaper. I will begin with the *Los Angeles Times*, which has a liberal-leaning. This time I would like to focus on the term 'white'. As can be seen from the figure below, the qualifying adjective 'white' is used only once out of 8 results to identify the race. If, on the other hand, one searches for the adjective 'black', there are 11 times as many sentences in which it has a qualifying function (88). The press uses expressions such as 'black man', 'black American', 'black activist', 'black people', 'black protesters', and so on. So why is it that in the case of an Afro-American person there is all this need to emphasize the color of the skin as well and when talking about a white person, this is practically never emphasized?

harbor in the distance. Notably, he's dressed in red, white and blue. Most everything else is from the 20th century, especially folk art. Some have criticized Mapplethorpe, a white artist, as fetishizing Black sexuality, but the omission is puzzling. Why? Assumptions about race. Dressed in a long but unstuffy white evening gown, the geometric decorations on its billowing skirt a nod to the architecture of St. John's Church across Lafayette Park from the White House, as they gathered to protest the death of George Floyd on June 5, the artist Healy worked as tensions around Reconstruction rose and fell. It was the summer before his much-publicized walk from the White House to a nearby church. U.S. District Judge Dabney Friedrich on Monday, hangs nearby. The unknown artist — most likely a white man, given the racial and gender limitations imposed around studio access — focused elegance. Her skin is gray, as in a black-and-white photograph, the artist's frequent strategy for asking viewers of a painting

Figure 15: Concordances of the term 'white'

Looking at the word frequency list I noticed that the personal name 'Abdullah' was repeated 43 times in this corpus, so this attracted my attention. I found out that Melina Abdullah is one of the co-founders of the BLM movement in Los Angeles.

In this corpus her name is cited in two articles. The first one talks about the death and violence threats received by the African-American leader, which have been classified by the police as actions of "swatting".⁸ Abdullah claims that these threats are the result of hatred and racism simply because she stands up for the rights of African-Americans (Rector, 2021, *Los Angeles Times*). However, this shows how little support and protection

⁸ Swatting: "Swatting occurs when someone falsely reports an emergency in order to draw a large police presence to a location. Such incidents are considered dangerous for the targets and have been deadly in the past." (<https://www.latimes.com/california/story/2021-09-29/black-lives-matter-melina-abdullah-third-swatting-incident>)

there is from the Los Angeles Police Department for a black leader. In the second article, the arrest of Abdullah for attempted violence against a police officer is discussed. After a thorough examination of the incident, Melina was later found innocent. Thus, in both cases, we have incidents of discrimination (such as lack of protection and police indifference) or false accusations of violence against the activist.

Curtis said Abdullah grabbed him by the arm to stop him from taking Hines-Brim in-to custody. claiming to be her young son alleged that Abdullah had overdosed on pills and needed help. less than two hours before returning its verdict that Abdullah was not wrongfully arrested by then-LAPD Det. Jason Curtis.

Figure 16: Concordance of the noun 'Abdullah'

I will now focus on the *New York Post* sub-corpus, which has a more conservative-leaning. I would like to start by searching for the adjective 'white'. In this sub-corpus it occurs 11 times. As can be seen in the figure below, the qualifying article is mostly used to describe people of white ethnicity in contrast to the results of the *Los Angeles Times*, where 'white' was used very little to identify people. In particular, I would like to focus on the last five results that talk about the concept of 'white supremacy'. Reading through the articles, I noticed that 'white supremacy' refers in all cases to the attack on Capitol Hill in January 2021, which we have talked about before.

by civil-rights activists over the violent riots outside the White House from May 29 through June 3, 2020. These "mostly peaceful 'white man in blackface, and a very conservative-minded white man at that." Black Lives Matter leader Hawk Newsome blasted Adams by using a racial slur and called him a "white man in blackface, and a very conservative-minded white man to return to their jobs. BLM's Hawk Newsome blames 'white man' Mayor Eric Adams for NYC 'war zone' <https://nypost.com>. Mayor Eric Adams of being "a very conservative-minded white man." William Farrington The number of victims was up 15.9 per cent stopping and arresting Black people at a higher rate than white people, more frequent use of force on people of color during his confirmation to "supervise the prosecution of white supremacists and others who stormed the Capitol on Jan. 6." In ittee keeps releasing incendiary public statements about "white supremacists," and Pelosi has declared Trump an "accessory" to m is against Trump supporters, who the Democrats cast as "white supremacists"? Biden began his presidency by hyping the Jan. 6 r defamed as "insurrectionists," "domestic terrorists" and "white supremacists." How are they supposed to get a fair trial democracy since the Civil War" and laying the blame on "white supremacy and domestic terrorism" which he claimed is "the mos

Figure 17: Concordance of the term 'white'

In 2022 to hear talk of 'white supremacy' again is disturbing. Along with the abolition of slavery in the US in 1865 with the passing of the Thirteenth Amendment to the US Constitution⁹, this retrograde concept should also have died out.

⁹ *Il Sole 24 Ore* (https://st.ilssole24ore.com/art/notizie/2014-12-29/150-anni-fa-18-dicembre--abolizione-schiavitu-stati-uniti-131953.shtml?uuiid=ABkrQUWC&refresh_ce=1)

The next research focuses on two people, who are mentioned several times in this sub-corpus: Eric Adams which occurs 30 times and Hawk Newsome which occurs 16 times. Eric Adams is New York's mayor and Hawk Newsome is the co-founder of Black Lives Matter in New York. In this sub-corpus, there are some articles that speak of events that occurred between these two leaders. In the figure below, one can see how the name 'Newsome' is followed by accusatory verbs against mayor Eric Adams. Reading the context of the article, it can be deduced what happened between the two, i.e. several threats. In the first place, the leader Newsome had threatened Adams with starting riots and fires in the city if the mayor hadn't stopped the increasing violence and shootings in the Big Apple as soon as possible. Moreover, the black leader "referred to Adams by using a racial slur and called him a 'white man in blackface', and a very conservative-minded white man at that. [...] Newsome blasted Mayor Eric Adams for being a "master of the press conference" at crime scenes."¹⁰. Analyzing the articles in depth, we can see how a leader who fights to abolish racism, proves to be racist at the same time.

fe. This city is a war zone and he can't stop it." Newsome, a co-founder of Black Lives Matter of Greater New York, also referred to Adams b
ler. During an appearance on Fox News, Hawk Newsome accused Adams of failing to protect black New Yorkers from surging gun violence
compared to the same period last year. Hawk Newsome accused Mayor Eric Adams of being "a very conservative-minded white man." Will
ing in a subway car in Brooklyn. On Thursday, Newsome also criticized Adams for calling out the silence of BLM and anti-cop activists follo
or people of color hardest. Which makes Hawk Newsome and his ilk the real bane on the black community — which, incidentally, voted ove
' <https://nypost.com/2022/04/21/blms-hawk-newsome-blames-mayor-eric-adams-for-nyc-war-zone/> Bruce Golding April 21, 2022 6:07
re allowed to return to their jobs. BLM's Hawk Newsome blames 'white man' Mayor Eric Adams for NYC 'war zone' <https://nypost.com/2022/04/21/blms-hawk-newsome-blames-mayor-eric-adams-for-nyc-war-zone/>
e man at that." Black Lives Matter leader Hawk Newsome blasted Mayor Eric Adams for being a "master of the press conference" at crime s
n urban homicides alone rising 5 percent. And Newsome & Co. are part of the problem. As the Manhattan Institute's Heather Mac Donald p
ed." Mayor Adams previously called out Hawk Newsome for threatening to hold riots in response to higher police measures. Lev Radin/Pac
athsome local Black Lives Matter leader Hawk Newsome just called Mayor Eric Adams a "c-n" live on national TV for pointing out BLM'

Figure 18: Concordances of the noun 'Newsome'

Regarding 'Adams' I would like to focus on one outcome in particular. To analyse it, however I have to quote the full sentence: "Eric Adams emphasized the victims of the shootings were black, but he feels that BLM has not given enough attention to the gun violence over the weekend. The mayor said those who protest against police brutality should also acknowledge the scourge of gun violence" (See footnote 3).

¹⁰ New York Post (<https://nypost.com/2022/04/21/blms-hawk-newsome-blames-mayor-eric-adams-for-nyc-war-zone/>)

Eric Adams emphasized the victims of the shootings were black, but he feels that BLM has not given enough

Figure 19: Concordance of the noun 'Adams'

In this statement, it seems that the mayor wants to blame the deaths of black people on black people themselves. By saying that the BLM has not paid enough attention to gun violence, he makes it clear how he is delegating responsibility for gun violence to black people. Their only fault, therefore, seems to be the color of their skin. Again still in 2022, we are forced to read statements full of racism and hatred that, with their mere words, are able to erase years of struggles and fights for equal rights and an end to racial discrimination.

3.3 Connotations, metaphor and hyperbole

After conducting this in-depth analysis and with the results obtained, I was able to deduce that generally the American press compared to the British one may tend to use negative connotations more often when describing ethnic minorities. The British press might seem, most of the time, to be more liberal and neutral in reporting this kind of news.¹¹ However, there are some cases where black people are negatively connoted. There are two ways of doing this: through metaphors, i.e. replacing the term with a figurative term, or hyperbole, i.e. providing an alteration of reality itself.

I will start with metaphors. The first term which can be considered as a metaphor is the verb 'to flood'. In the figure below we can see that the crowd of local residents who were participating in a protest over the death of Michael Brown, is compared to a natural element: water. In this sentence, we can see the combination of hyperbole and metaphor to emphasize the large numbers of residents in the streets.

Hundreds, and then thousands, of local residents had flooded the streets. For the Ferguson

Figure 20: 'Flooded' as metaphor

¹¹ USC Center on Public Diplomacy (<https://uscpublicdiplomacy.org/blog/mainstream-media-discrepancies-between-us-and-uk>)

The second term used as a metaphor is the noun 'fight'. In the general corpus this noun occurs 15 times and in the figure below I have reported some of the outcomes. As we can see most of the time it is used as a verb and in a few cases as a noun. The verb 'to fight' is always used as a metaphor, since we are not talking about an actual armed battle on a war field, but rather about a moral, symbolic and justice battle. In fact, we notice how 'fight' is followed by terms such as 'for black people/lives', 'for racial justice', 'for their country'.

or a reason — to fight for Black lives.” Despite the wave c
 hat people do to fight for black people, and this is what t
 ; attention to the fight for civil rights and voting rights. Tl
 more broadly the fight for human rights. The words “I car
 one another and fight for Justice for Jim Rogers and for
 : It is our duty to fight for racial justice. It is our duty to w
 : It is our duty to fight for racial justice. It is our duty to w
 y were taught to fight for their country against any enem
 said, they didn’t fight it. They waited for it to pass, believ
 de movement to fight racial injustice. “They have had a tr

Figure 21: ‘To fight’ as metaphor

Continuing with hyperboles we can start with the term ‘huge’, which according to Cambridge Dictionary means “extremely large in size or amount” (Cambridge Dictionary). Searching this adjective in the corpus I found the sentence “A huge crowd of protesters has gathered in London to throw their support behind the Black Lives Matter movement [...]”. ‘Huge’ may have both positive and negative connotations, but in this case, it seems more negative. In this context ‘huge’ is used to highlight the fact that there were so many people at the demonstration that the protest may have been unsafe for the community. Words indicating quantity are often used as hyperbole. In fact, there are several hits with the word 'hundreds' in the corpus. As can be seen in the sentences below, ‘hundreds’ is used to denote a large group of protestors that invade the streets.

Hundreds join march to protest against systemic racism in the UK Demonstrators take to streets
 Hundreds of demonstrators took to the streets of west London on Sunday in the first ever

Figure 22: ‘Hundreds’ as hyperbole

In the example below, instead, one can find the term ‘hundreds’ followed by ‘of peaceful demonstrators’, so it has a positive connotation because the protesters are highlighted as being ‘peaceful’.

hundreds of peaceful demonstrators amid protests over the police murder of George Floyd

Figure 23: ‘Hundreds’ as hyperbole

I want to analyze this term in-depth and see in what contexts it is used in the general corpus. First of all, ‘peaceful’ occurs 26 times. As we can see from the outcomes, most of the time the adjective ‘peaceful’ is followed by ‘protest’ or ‘protesters’. We can therefore deduce that the protests carried out by the Black Lives Matter movement are mainly peaceful protests, without violent or rebellious aims. We can say that, considering the general trend of the analyses conducted so far, these are unexpected outcomes, as most of the terms analyzed in this dissertation had a discriminatory tendency.

id-degree murder. Following an afternoon of peaceful demonstrations in the park, including an impassioned speech by Star Wars ac
cement officers forcibly cleared hundreds of peaceful demonstrators amid protests over the police murder of George Floyd shortly
orate the same spirit of freedom through peaceful protest. About 400 demonstrators walked along Bayswater Road from Notting
:" charges that came as a result of acts of peaceful protest. I was the first journalist to end up in cuffs while covering the unrest. Re
Mr. Brown wants to be able to engage in peaceful protest in the City in the future, without concern or fear that he will be arrested
are accused of using force to break up a peaceful protest. "Reasonable officers would have known that such alleged actions viola
441/ Up to 5,000 people took part in the peaceful protest today, Sussex Police said (Picture: PA) Around 5,000 people have march
441/ Up to 5,000 people took part in the peaceful protest today, Sussex Police said (Picture: PA) Around 5,000 people have march
PD officers rushed at and knocked down peaceful protesters assembled in front of the Mayor's mansion and then brutally beat th
waters in a very real way that will allow peaceful protesters to get swept up in arrests," Florida Rep. Ben Diamond (D) said at the
that the department unjustly focused on peaceful protesters while ignoring individuals who were burglarizing stores and commit
29 through June 3, 2020. These "mostly peaceful protests," as The Washington Post describes them, were so violent that then-Pri
nd to train officers in lawful responses to peaceful protests." On anniversary of George Floyd's death, L.A. rally declares, 'Black live

Figure 24: Concordances of the term ‘peaceful’

If we look again for a quantitative adjective like ‘thousands’, we obtain 30 results. Most of them concern the quantities of protestors. Some examples are: “Thousands headed to the US Embassy in Battersea”; “Thousands gathered in Hyde Park on Wednesday calling for justice”; “Thousands of protesters were arrested in the months following George Floyd’s death”. Another hyperbole I noticed in the general corpus is the adjective ‘sweeping’, which according to Cambridge Dictionary means “affecting many people, large” (Cambridge Dictionary). As can be seen in the figure below, this adjective is used to emphasize and magnify situations such as protests or restrictions.

residents could soon face under sweeping anti-protest legislation championed by C
ry, local nonprofits have brought sweeping legal overhauls and incremental change
sure people in Florida don't face sweeping re-strictions to employment, housing an
e most detailed picture yet of the sweeping scope of NYPD surveillance during mass

Figure 25: 'Sweeping as hyperbole'

Conclusion

This last chapter is the most important one because in this linguistic analysis I have tried to touch on as many aspects as possible covered in the first two chapters. The analysis tool AntConc was fundamental in my study. Initially, I introduced the analysis by talking about the word frequency of both the general corpus and the various sub-corpora concerning each newspaper.

For each newspaper, I carefully selected specific articles that dealt with the topic of racial discrimination, especially regarding the Black Lives Matter movement. The analysis then continued by focusing on collocation and concordance. I mainly focused on words related to the topic of racism against black people and compared the results I obtained in the general corpus with those obtained in the sub-corpora. I was able to derive results that arose reasoning in me, sometimes personal, sometimes related to the general context. The last section of the chapter deals with the different connotations that certain words can have depending on the context in which they are used and also the rhetorical figures (such as metaphors and hyperbole) that are used to alter or emphasize the meaning of the word.

Conclusion

The main topic of this dissertation was racial discrimination against Afro-Americans from a linguistic perspective. The chapters have been placed in chronological order so that the reader can acquire the basic knowledge of the topic, which is then applied in the final analysis. Thanks to this work, I was able to observe the impact of the media in the dissemination of racism and note how very often representations of minorities are highlighted in a negative light. I was also able to observe how the reception of news can be completely altered through linguistic techniques and through the creation of negative mental models. I realized that the discrimination of non-white journalists can be interpreted as a declination of the colonial past and the slave narrative, in fact, I gave the example of Frederick Douglass and his narrative of the racial discrimination he experienced on his skin on the plantations.

Thanks to this study, I was able to gain an in-depth understanding of how the media try to shape our thoughts and opinions to their liking. In the analysis I carried out, I wanted to focus both on terms that recurred most often in the corpus, and on terms that attracted my attention, such as personal names that allowed me to investigate the events in depth. In fact, from the results, I was able to deduce that in 90% of the cases they were episodes of violence and racial discrimination. In this work, I wanted to compare the British press with the US press and contrast different newspapers and tabloids that had contrasting political orientations. The results obtained were sometimes very predictable, but sometimes totally unexpected. In general, I was able to deduce that the representation of minorities in the media is very often negatively compromised, precisely because, as I explained in the first chapter, there is a large predominance of white journalists who therefore avoid putting their own ethnicity in a bad light.

The articles I selected all dealt with episodes relating to the struggle of the Black Lives Matter movement, yet from the results, I was able to deduce how the figure of the police is always present, so this was another confirmation of how these struggles are depicted

as violent and dangerous for the community. In this analysis, I was also able to see how certain terms can be used with completely different connotations, depending on the context in which they are used.

As I have argued before, language is one of the most powerful weapons we possess. In my view, it is precisely through this powerful tool that we should begin to raise awareness on the issue of racial discrimination, starting with the major sources of information: the media. As far as I see things, I believe that language studies should look deeper into this issue and understand why, in the 21st century, the phenomenon of racism has not yet died out. The concept of 'white supremacy' should no longer exist in the media, nor socially, culturally or ethnically. In an era where racial discrimination is no longer a tolerable phenomenon, one should start dismantling it from its roots: starting with its verbal dissemination and through language. It is my hope that this inequality, which is still so glaring today, will be alleviated as the years and struggles go by, and that the voices of the forgotten and oppressed will finally be heard, to finally provide the justice and equal rights that everyone deserves, regardless of race, social class, gender or sexual orientation.

Bibliography

- Anthony, L. 2013. A critical look at software tools in corpus linguistics. *Waseda University*. 141-161.
- Baker, P., 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society*. 273-306.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Bloomsbury Publishing.
- Biber, D., Conrad, S., Reppen, R. 1998. *Corpus linguistics: investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., Leech, G. 1999. *Longman Grammar of Spoken English*. London: Longman.
- Biber, D. 2009. A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*. 275-311.
- Busà, M. Grazia. 2013. *Introducing the Language of the News: A student's guide*. London: Routledge.
- Chomsky, N., Herman, E. 1988. *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.
- Cook, G. 1989, *Discourse*. Oxford: Oxford University Press.
- Fowler, R. 1991. *Language in the News. Discourse and Ideology in the Press*. London: Routledge.
- Mooney, A., Evans, B. 2015. *Language and the Media*. In A. Mooney & B. Evans (eds.) *Language, Society and Power*. London: Routledge. 69-98.
- McEnery, A., Wilson, A. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Partington, A. 2008. The Armchair and the Machine: Corpus-assisted Discourse Research. In C. Torsello, K. Ackerley, E. Castello (eds.) *Corpora for University Language Teachers*. Bern: Peter Lang. 95-117.
- Reppen R., Simpson-Vlach, R. 2020. Corpus linguistics. In N. Schmitt, M. Rodgers. *An Introduction to Applied Linguistics*. London: Routledge. 91-108.
- Sinclair, J. 2004. *Trust the text: Language, Corpus and Discourse*. London: Routledge.
- Tognini-Bonelli, 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins Publishing Company.
- Taylor, Stephanie, 2014. *Discourse analysis*. In Teo, Thomas ed. *Encyclopedia of Critical Psychology*. New York: Springer. 449-457.

- Van Dijk, T. n.d. New(s) Racism: A Discourse Analytical Approach. In S. Cottle (ed.) *Ethnic Minorities and the Media*. Philadelphia: Open University Press. 33-49.
- Van Dijk, T. n.d. Mediating racism: the role of the media in the reproduction of racism. *Language, power and ideology*. Amsterdam: University of Amsterdam. 199-226.
- Van Dijk, T. 2006. Racism and the European Press. *Presentation for the European Commission against Racism and Intolerance*. Strasbourg. 1-20.
- Van Dijk, T. 2015. Racism in the Press. In N. Bonvillain (ed.), *The Routledge Handbook of Linguistic Anthropology*. London: Routledge. 384-392.
- Van Dijk, T. 1991. *Racism and the press*. London: Routledge.
- Weir, G., Anagnostou, N. 2007. Exploring newspapers: a case study in corpus analysis. *ICTATLL Workshop*. 1-8.
- O'Keeffe, A., McCarthy, M. 2010. *The Routledge Handbook of Corpus linguistics*. London: Routledge.
- Willnat, L., Weaver, D., Wilhoit, G. 2017. *The American Journalist in the Digital Age: A Half-Century Perspective*. New York: Peter Lang.
- Zapata Barrero, R., Yalaz, E. 2018. Discourse and Migration. In B. Zapata Barrero, E. Yalaz, (eds.) *Qualitative Research in European Migration Studies*. New York: SpringerOpen. 227-245.

Websites

- Craig, T. 2021. He was arrested at a Black Lives Matter protest. Now, he's warning others about Florida's anti-riot proposal. https://www.washingtonpost.com/national/florida-protest-bill-silences-activists/2021/04/09/5b5b252c-93c2-11eb-a74e-1f4cf89fd948_story.html%20CAP%203.1 (Accessed 2 May 2022)
- Differenza tra bottom up e top down? n.d. <https://qualcherisposta.it/differenza-tra-bottom-up-e-top-down> (Accessed 8 May 2022)
- Engel, V. 2019. Modern Racism <https://www.opendemocracy.net/en/can-europe-make-it/modern-racism/> (Accessed 15 April 2022)
- EJI, 2021. FBI Reports Hate Crimes at Highest Level in 12 Years. <https://eji.org/news/fbi-reports-hate-crimes-at-highest-level-in-12-years/> (Accessed 7 April 2022)
- Fornara, P. 2015. I 24 anniversari del 2015: dalla Magna Charta del re d'Inghilterra fino alla Merkel. https://st.ilsole24ore.com/art/notizie/2014-12-29/150-anni-fa-18-dicembre--abolizione-schiavitu-stati-uniti-131953.shtml?uuid=ABkrQUWC&refresh_ce=1 (Accessed 18 May 2022)

- FutureLearn, n.d. What is a corpus? <https://www.futurelearn.com/info/courses/understanding-dictionaries/0/steps/62809> (Accessed 13 May 2022)
- Gee, K. 2018. MAINSTREAM MEDIA DISCREPANCIES BETWEEN THE U.S. AND THE UK <https://uscpublicdiplomacy.org/blog/mainstream-media-discrepancies-between-us-and-uk> (Accessed 15 May 2022)
- Gli approcci corpus based e corpus driven, n.d. <https://123dok.org/article/gli-approcci-corpus-based-e-corpus-driven.dzxejwvq> (Accessed 10 May 2022)
- Goldberg, Z. 2020. How the Media Led the Great Racial Awakening <https://www.tabletmag.com/sections/news/articles/media-great-racial-awakening> (Accessed 1 April 2022)
- Hogan, B., Golding, B. 2022. BLM's Hawk Newsome blames 'white man' Mayor Eric Adams for NYC 'war zone'. <https://nypost.com/2022/04/21/blms-hawk-newsome-blames-mayor-eric-adams-for-nyc-war-zone/> (Accessed 18 May 2022)
- Huge. n.d. Cambridge Dictionary. Cambridge University Press. <https://dictionary.cambridge.org/it/dizionario/inglese/huge> (Accessed 5 June 2022)
- Lowery, W. 2017. Black Lives Matter: Birth of a movement. <https://www.theguardian.com/us-news/2017/jan/17/black-lives-matter-birth-of-a-movement> (Accessed 10 May 2022)
- Mi Score, n.d. https://www.sketchengine.eu/my_keywords/mi-score/#:~:text=The%20Mutual%20Information%20score%20expresses,score%20which%20may%20be%20misleading (Accessed 17 May 2022)
- Qualitative vs Quantitative Analysis, n.d. <http://www2.sal.tohoku.ac.jp/ling/corpus3/3qual.htm> (Accessed 10 May 2022)
- Rector, K. 2021. Black Lives Matter leader targeted in third 'swatting' incident, the second since last week. <https://www.latimes.com/california/story/2021-09-29/black-lives-matter-melina-abdullah-third-swatting-incident> (Accessed 3 May 2022)
- Rozado, D., Al-Gharbi, M., Halberstadt, J. 2022. Use of 'sexist' and 'racist' in the New York Times increased over 400% since 2012. Why? <https://www.theguardian.com/commentisfree/2022/feb/26/media-news-article-shift-discourse-language> (Accessed 3 April 2022)
- Smith, M. 2017. How Left or right-wing are the UK's newspapers? <https://yougov.co.uk/topics/politics/articles-reports/2017/03/07/how-left-or-right-wing-are-uks-newspapers> (Accessed 3 April 2022)

Sweeping. n.d. Cambridge Dictionary. Cambridge University Press.
<https://dictionary.cambridge.org/it/dizionario/inglese/sweeping> (Accessed
1 June 2022)

Wikipedia, the Free Encyclopedia, n.d., Taking the knee
[https://en.wikipedia.org/w/index.php?title=Taking the knee&oldid=10916
98715](https://en.wikipedia.org/w/index.php?title=Taking_the_knee&oldid=1091698715) (Accessed 8 May 2022)

Riassunto in italiano

Il mondo sta facendo progressi in molti settori, dalla tecnologia alla sostenibilità. Tuttavia, quando si tratta di discriminazione razziale, non possiamo dire lo stesso. Viviamo in un'epoca in cui la "diversità", sia essa etnica, razziale, di genere o di orientamento sessuale, è ancora vista come una minaccia, un problema da risolvere, quando in realtà è la diversità stessa a rendere le città, le culture, gli ambienti e gli spazi in cui viviamo ricchi. Poiché la lingua è lo strumento che ci permette di parlare, di esprimerci e di sentirci liberi, ho voluto dedicare la mia tesi di laurea al tema della discriminazione razziale, trattato però da una prospettiva linguistica. L'area geografica di interesse è quella anglofona, infatti l'analisi finale della tesi si concentra sul movimento attivista per i diritti degli afroamericani: il *Black Lives Matter*.

L'obiettivo principale della mia tesi è portare alla luce il modo in cui i media britannici e americani riportano le notizie e i fatti relativi alle lotte per l'uguaglianza razziale ed evidenziarne i risultati attraverso un'analisi linguistica. La mia tesi è composta da tre capitoli. L'ordine degli argomenti trattati in ogni capitolo è studiato apposta per fornire al lettore dapprima le conoscenze e i concetti generali sull'influenza dei media e su come si svolge un'analisi linguistica, per poi concludere con l'applicazione di tali concetti nell'analisi vera e propria.

Nel primo capitolo viene fornita una panoramica del ruolo svolto dai media nella rappresentazione delle minoranze etniche e di come contribuiscono alla diffusione del razzismo. Vedremo inoltre come le scelte linguistiche e le tecniche utilizzate siano in grado di plasmare l'interpretazione di ciò che viene descritto. In questo capitolo ho voluto dedicare una sezione alla discriminazione subita dai

giornalisti non bianchi e la loro difficoltà nel far sentire la propria voce in un mondo in cui prevale la supremazia bianca.

Il secondo capitolo approfondisce l'aspetto teorico dell'analisi linguistica dei corpora e si concentra maggiormente sugli aspetti principali della realizzazione di un'analisi linguistica. Viene fornita un'introduzione su cosa sia un corpus, su come si costruisca e su quali siano le dimensioni ideali. Verso la fine del secondo capitolo, c'è un approccio graduale al terzo capitolo, poiché vengono introdotti gli aspetti della frequenza, della concordanza e della collocazione, che costituiscono la parte centrale dell'analisi linguistica del terzo capitolo.

Il terzo capitolo può essere considerato il più rilevante perché in esso vi sono inclusi tutti i concetti menzionati nei primi due capitoli. Questo capitolo è interamente dedicato all'analisi di un corpus e di diversi subcorpora creati da me, composti interamente da articoli riguardanti il movimento *Black Lives Matter*. Mi sono concentrata su termini che hanno un chiaro legame con la discriminazione razziale, per vedere e analizzare i contesti in cui vengono utilizzati. L'analisi si concentra principalmente sugli aspetti della frequenza, la concordanza, la collocazione e sulle diverse connotazioni che specifici termini possono avere in base al contesto d'uso.

Con questa tesi mi è stato possibile comprendere a fondo l'impatto dei media nella diffusione del razzismo e ho potuto notare come molto spesso le rappresentazioni delle minoranze siano evidenziate in modo negativo, perché i media cercano spesso di plasmare i nostri pensieri e le nostre opinioni a loro piacimento. Nell'analisi che ho svolto, ho voluto concentrarmi sia sui termini che ricorrevano più spesso nel corpus, sia su quelli che hanno attirato la mia attenzione, come nomi di persone che mi hanno permesso di approfondire determinati eventi di violenza. Dai risultati, infatti, ho potuto dedurre che nel

90% dei casi si trattava di episodi di violenza e discriminazione razziale. In questo lavoro ho voluto confrontare la stampa britannica con quella statunitense e contrapporre diversi giornali e tabloid con orientamenti politici contrastanti. I risultati ottenuti sono stati a volte molto prevedibili, ma a volte del tutto inaspettati. In generale, ho potuto dedurre che la rappresentazione delle minoranze nei media è molto spesso compromessa negativamente, proprio perché, come ho spiegato nel primo capitolo, c'è una grande prevalenza di giornalisti bianchi che quindi evitano di mettere in cattiva luce la propria etnia.

Gli articoli selezionati trattavano tutti episodi relativi alle lotte del movimento *Black Lives Matter*, eppure dai risultati ho potuto dedurre come la figura della polizia sia sempre presente, quindi questa è stata un'ulteriore conferma di come queste lotte vengano dipinte come violente e pericolose per la comunità.

Come ho sostenuto in precedenza, il linguaggio è una delle armi più potenti che possediamo. A mio avviso, è proprio attraverso questo potente strumento che dovremmo iniziare a sensibilizzare l'opinione pubblica sul tema della discriminazione razziale, a partire dalle principali fonti di informazione: i media. Credo che gli studi linguistici debbano approfondire la questione razziale e capire perché, nel XXI secolo, il fenomeno del razzismo non si sia ancora estinto. Il concetto di "supremazia bianca" non dovrebbe più esistere nei media, né a livello sociale, culturale o etnico. In un'epoca in cui la discriminazione razziale non è più un fenomeno tollerabile, si dovrebbe iniziare a smantellarla dalle sue radici: a partire dalla sua diffusione verbale e attraverso il linguaggio. La mia speranza è che questa disuguaglianza, ancora oggi così evidente, si attenui con il passare degli anni e delle lotte, e che le voci dei dimenticati e degli oppressi vengano finalmente ascoltate, per rendere finalmente la giustizia e la parità di diritti che tutti meritano, indipendentemente dalla razza, dalla classe sociale, dal genere o dall'orientamento sessuale.