A CONCEPTUALISATION OF OLYMPIC LEGACY DISCOURSE IN BRITISH AND BRAZILIAN MEDIA COVERAGE OF THE LONDON 2012 AND RIO 2016 SUMMER OLYMPIC GAMES

Caio de Castro Mello Santos

School of Advanced Study

University of London

July 2025

I, Caio de Castro Mello Santos confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

I would like to thank my supervisors Professor Jane Winters and Martin Steer for the support provided, which made this thesis possible. Besides their technical competence, expertise and professionalism, both of them were able to see the human behind the student. I must highlight their generosity in listening to me, reading my work, providing careful and detailed feedback, and devoting their time and resources to my learning experience. I feel extremely grateful to have worked in such an encouraging and constructive environment, where I could have material and immaterial resources to develop my full potential. This, unfortunately rare, positive and healthy academic environment is a product of Professor Winters' leadership and her brilliant team in the Digital Humanities Research Hub at the School of Advanced Study. Therefore, I must extend my thanks to the members of the team Dr. Christopher Ohge, Dr. Gabriel Bodard, Dr. Naomi Wells, Kunika Kono, Valerie James, Michael Donnay, Sam Sakhabuth, Niilante Ogunsola-Ribeiro, Simon Parr and Dr. Megan Bushnell. Special thanks to Elaine Walters, Director of the School of Advanced Study, for being so attentive to the students' needs, always interested in improving our experience and creating space for personal development.

I thank my colleagues in the PhD programme at SAS Francielle Carpenedo, Rhiannon Lewis and Federico Filauri for making me feel part of a community. In particular, I would like to thank my wonderful colleague Daniela Major, with whom I have shared the entire journey, both the excitement of discovering something new and the frustrations that are part of a thesis writing process.

The study presented in this thesis was developed over unprecedented times, during the Covid-19 pandemic. This means that despite the inherent challenges of a PhD, a significant part of my work had to be done with restricted access to resources, locked in a small flat in London and in a situation of extreme social solitude. If not for all the people above

mentioned, I would not have been able to keep moving forward and this work would not have been possible.

Over the course of the PhD, I have had the honour to work with fantastic people, who I thank very much for having taken the time to have conversations about my project and sharing their knowledge with me. I would like to thank the UK Web Archive team at the British Library, in particular Jason Webber, Helena Byrne, Carlos Rarugal and Ian Cooke. Also the Programming Historian team, especially Adam Crymble, Alex Wermer-Colan, Anisa Hawes, Daniel Alves and Jimmy Medeiros, for providing me with a space to learn, which was fundamental for developing my coding skills, as well as for establishing a network that gave me the opportunity to be involved in several activities such as building learning material, translating and editing content.

This thesis was made possible thanks to the CLEOPATRA project, funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant. Therefore, I would like to thank the European Union and their taxpayers for investing in science and providing me with the opportunity to study and develop knowledge. I must also mention all the professors involved, the dedicated project manager Dr. Ivana Marenzi and the wonderful colleagues at CLEOPATRA with whom I established collaborations and friendship: Gaurish Thakkar, Gullal Singh Cheema, Gabriel Amaral, Tin Kuculo, Sara Abdollahi, Elisavet Koutsiana, Jason Armitage, Endri Kacupaj, Golsa Tahmasebzadeh, Daniela Major, Alberto Olivieri, Abdul Sittar, Swati, Sahar Tahmasebi, Anna Katrine Jørgensen and Diego Alves.

My academic life has been the result of public investment, which supported me to reach the point of being a student at the University of London. Therefore, I would also like to extend my thanks to the Brazilian funding agencies, namely the Foundation for the Support of Science and Technology in Pernambuco (FACEPE) and the Brazilian National Council for Scientific and Technological Development (CNPq) for funding my undergraduate and Master's research, as well as the German Ministerium für Kultur und Wissenschaft des

Abstract

The Summer Olympic Games are held every four years in a different city around the globe, promoting what is considered one of the biggest sporting events in the world. For more than a century of existence, the modern Olympiad has been demanding exorbitant investments from local governments for the development of the infrastructure required to adapt host cities to accommodate the competition as well as athletes and visitors. For this reason, the event has faced criticism for the constant increase of cost that has often been seen as inappropriate by local residents. In this context, the Olympic legacy – what is constructed or developed for the Games – has become the central strategy of the International Olympic Committee to justify the use of scarce public resources. The concept however is disputed and appropriated by different social and political actors, with its meaning not being strictly defined.

This thesis investigates how the Olympic legacy discourse has been framed by the British and Brazilian media coverage of the London 2012 and Rio 2016 Summer Olympic Games. The corpus is composed of news articles collected from both the live web and web archives, published by British and Brazilian media outlets, in the English and Portuguese languages. The discourse analysis of the material was conducted using distant reading techniques, based on Natural Language Processing (NLP) tools such as tokenization, word counts, entity recognition, sentiment analysis and network visualisations. These methods were combined with a qualitative close reading of the texts, informed by the discourse analysis tradition. The main objective of this thesis is to provide a conceptualisation of the word 'legacy' based on the news media narratives, as well as on the 'silences' of those narratives. The thesis contributes to a wider and more comprehensive understanding of the multiple players involved and the tension provoked by their voices in the construction of an 'Olympic legacy'.

Contents

Abbreviations

AGLO: Rio Olympic Legacy Governance Authority

AI: Artificial Intelligence

ANT: Actor-Network Theory

APO: Public Olympic Authority

BERT: Bidirectional Encoder Representations from Transformers

BRICS: Brazil, Russia, India, China and South Africa

BRT: Bus Rapid Transport

CLEOPATRA: Cross-lingual Event-centric Open Analytics Research Academy

ComCat: Catalysing Communities

CPLP: Community of Portuguese Language Countries

DA: Discourse Analysis

DCMS: Department for Digital, Culture, Media and Sport

DH: Digital Humanities

EOM: Municipal Olympic Company

EU: European Union

GEO: Experimental Olympic Gymnasium

IOC: International Olympic Committee

IPC: International Paralympic Committee

LBCD: Brazilian Anti-doping Lab

MP: Minister of Parliament

MPF: Ministério Público Federal (Brazilian Public Prosecution Service)

NER: Named Entity Recognition

NLP: Natural Language Processing

NLTK: Natural Language Toolkit

Ofcom: Office of Communications

PPP: Public-Private Partnerships

PT: Workers Party

SA: Sentiment Analysis

TF-IDF: Term Frequency – Inverse Document Frequency

UK: United Kingdom

UKWA: UK Web Archive

UNICEF: United Nations Children's Fund

URL: Uniform Resource Locator

List of figures

List of tables

Chapter 1: Introduction

The Summer Olympic Games are held every four years in a different city around the globe, promoting what is considered one of the biggest sporting events in the world. This four-year gap goes back to its origin in Ancient Greece, around 776 BC (Kanin, 2019), where it was an important temporal mark: instead of using years, time would have been counted as Olympiads. As pointed out by Kanin (2019), the ancient Olympic Games were used to demonstrate political power. The idea of turning the event into a competition with the potential for transforming people and promoting human moral development was only later introduced in the modern Olympic Games by Pierre de Coubertin in 1896. This idea was summarised in his quote, which later became the Olympic creed:[1] 'The important thing in life is not the triumph but the fight; the essential thing is not to have won, but to have fought well'.

Baron de Coubertin was a French aristocrat, historian and founder of the International Olympic Committee (IOC). He was responsible for designing the modern Olympic Games, its five rings symbol, as well as the values of Olympism. The latter was based on his interest in developing an educational project for France, with sports as pillars (The Olympic Studies Centre, 2022, p. 7).

Understanding the values of Olympism, promoted by Coubertin, is crucial for comprehending how and why the legacy of the Olympic Games has become an important topic to be defended and spread by the IOC as well as studied by scholars from several disciplines. Coubertin's ideas were, however, developed based on a romanticised idea of ancient Greece. As pointed out by Hill (1992, p. 6), Coubertin believed that the Greeks were the best model of a successful education system, due to the way it had gathered at the *gymnasia* people of different 'types' and generations and with multidisciplinary knowledge. He also believed that

---

[1] The Olympic creed as defined by the IOC on the webpage <https://olympics.com/ioc/faq/olympic-symbol-and-identity/what-is-the-olympic-creed>. Accessed on 12th April 2023.

the ancient Olympic Games were the highest level of sporting performance and, therefore, the reference to be aspired to.

Although this idea of using the event to promote social change was part of the discourse from the beginning of the modern era of the Games, it took some time until it was materialised in the word legacy. Legacy was mentioned for the first time in the 1956 Melbourne Olympic Games, but it was only in 1987 that the IOC organised a discussion around the topic (Preuss, 2019).[2] For Girginov (2018), it was only in the 1990s that hosting the event became connected to what he called the 'tangible measure of the three forms of cultural capital': social, economic and sporting benefits (2018, p. 3). As pointed out by Girginov and Preuss (2022, p. 2) 'the focus of Olympic aspirations has evolved over time and followed a progression from educating people through sport (i.e. micro) to addressing wider social and political objectives (i.e. macro) where sport was the driving force behind Olympism'.

Over time, the failure of host cities in guaranteeing a positive legacy for their citizens has threatened significantly the continuation of the Games, which had previously been taken for granted. Recently, cities like Boston, Budapest, Hamburg, Rome and Stockholm have given up on their candidacy to host the 2024 Summer Games, for example. This is despite the Olympic Movement's campaign to validate the rise in costs of the event (Karadakis, Bopp and Megan, 2019, p. 74). Despite its importance for guaranteeing the future of the event as well as its implications for the lives of citizens of host cities, no satisfying definition of legacy has been developed or agreed upon (Preuss, 2006).

Aiming at contributing to this gap in the understanding and definition of the word legacy, this thesis investigates the media coverage of the legacy of two events: the London 2012 and Rio 2016 Summer Olympic Games. These events were chosen for analysis as they were the two most recent Summer Games at the time when this thesis started to be developed in 2019.

---

[2] In his article 'Event legacy framework and measurement', Preuss (2019, p. 104) presented a thorough literature review of the emergence and transformations of the word legacy in the context of Olympism. The author highlighted Atlanta's mission statement goal of leaving a positive physical and spiritual legacy as a result of the Games. This idea of a spiritual legacy is imbricated and inspired by the values of Olympism, of an event aimed at transforming humankind.

They are also significant for comparison as they are Olympics that took place in very different circumstances: the United Kingdom was hosting the Games for the third time in history, while Rio 2016 was not only the first Games hosted in Brazil but also the first ever hosted in the Global South. In the next section, a brief contextualisation of both events will be provided.

1.1 London 2012: the third time as host

London has a long history with the Olympic Games, having hosted it for the first time in 1908. This Fourth Olympiad was pioneering in many aspects of mediatisation: it was the first to be filmed and to produce significant photographic coverage, which culminated in substantial media attention (Jenkins, 2011). According to Jenkins (2011), this documentation happened by chance: photographers working to document the Franco-British Exhibition, hosted in the London Olympic Stadium, ended up filming and photographing the sporting event happening in parallel in the same venue.

Over time, hosting the Olympics has gained more and more political importance. With Adolf Hitler's use of the 1936 Berlin Games to propagate his Nazi campaign, it became impractical to divorce the event from its communicative power. With the intention of celebrating the 2,600th anniversary of the Japanese Empire – and with the support of Germany and Italy – Tokyo was elected host city in 1940. Both the 1940 and 1944 Games, however, had to be cancelled due to the World War. The Games would only return in 1948, when London hosted the event for the second time. According to Hampton (2012), one of the main reasons why England had demonstrated interest in hosting the Games was to accelerate its economy after the war, by benefiting from ticket sales, tourism and the selling of British goods.

But why would London be interested in hosting the Summer Olympic Games for the third time in 2012? According to Poynter and MacRury (2012), London 2012, in a similar way to Beijing 2008, emerged in counterpoint to an economic model that privileged capital above politics, and resulted in inequalities and frustrations in civil society. For the authors, it was

part of a cultural strategy to respond to multiple crises derived from this system that threatened important societal achievements, such as public housing policies. In this scenario, the choice of East London to host the event, the most deprived area in the city, was of particular significance, as it reinforced the event as an opportunity to 'adjust' the problems caused by the recent decades of political and economic struggle. The event was, however, as will be evidenced throughout this thesis, a product and a producer of this neoliberal state itself.

The aim of using the Games as part of a broader cultural strategy is evidenced by the narrative construction of the opening ceremony. Thomas and Anthony (2015) analysed the British newspapers' commentary on the ceremony and what they called the 'competing constructions of British national identity'. According to the authors the ceremony was characterised by a celebration of socialist economic policies, represented by the National Health Service (NHS). For the commentators, 'the ceremony was a blatant and bold exercise in political theatre, alternately a rebellious statement against austerity and privatisation and a dangerous celebration of a misguided political doctrine' (Thomas and Anthony, p. 498), which gained particular meaning by being performed during the government of David Cameron, while he watched in the stadium's audience. Voices against the NHS were, of course, found within the newspapers' discourses, attesting to the controversies that form the debates as well as the role of the media as a site of tension and discursive dispute. These are two important concepts that run through this entire thesis.

While London used the Olympics in 2012 to reimagine its national identity and respond to social claims resulting from a neoliberal state, the subsequent Summer Olympics was framed as a celebration of decades of reduction in inequality, leading to a desirable promising future for the Global South.

1.2 Rio 2016: the future of the rocket nation

It was September 2009 when the magazine The Economist published its cover page entitled: 'Brazil takes off'.[3] The image on the cover page was of the famous Christ the Redeemer statue spitting fire from the base, as if it were a rocket about to fly. It was a time of celebration as Brazil's economy had gone through the global economic crisis of 2007-2008 with unexpected stability, had strengthened its partnership with the BRIC nations,[4] had gained prestige in the MERCOSUL, had expanded its middle class to include 29 million more people[5,6] and had become the leader in public policies against hunger among developing countries.[7] 2009 is also the year in which president Lula da Silva travelled to Copenhagen, in October, to be informed that Rio had won the bid and would host the 2016 Summer Olympic Games.

The award followed the 2007 Pan-American Games hosted in Rio and the 2007 election of Brazil as host of the 2014 FIFA World Cup, scheduling a sequence of major sporting events to be hosted by the country. President Lula da Silva's effusive celebration broadcast worldwide summarises the meaning of hosting the Games for the country: it was an opportunity to shed light on its recent fast growth as well as to place the nation among the leaders of the global agenda. Brazil was about to become the first country in the Global South to host the Olympic Games.

---

[3] Available at <https://www.economist.com/leaders/2009/11/12/brazil-takes-off>. Accessed on 22nd April 2023.

[4] The group of emergent fast-growing economies formed by Brazil, Russia, India and China. In 2010, it included South Africa and became the BRICS.

[5] 29 million people represented half of the population within this social-economic category.

[6] Available at <https://veja.abril.com.br/economia/mais-de-29-milhoes-entraram-para-classe-media>. Accessed on 23rd April 2023.

[7] Available at <https://www.bbc.com/portuguese/noticias/2009/10/091016_brasil_fome_relatorio_np>. Accessed on 23rd April 2023.

In September 2013, The Economist revisited its cover page from 2009. However, this time the rocket-statue flew uncontrolled, falling down from the sky, with its head facing down. This edition's title was: 'Has Brazil blown it?'. In that year, mass protests took to the streets of several Brazilian cities, the government of president Dilma Rousseff – Lula's successor – was facing a boycott by parliament, the economy was decelerating, and the optimistic wave had gone. President Rousseff was impeached in April 2016 and in August that year, the opening ceremony of the XXXI Olympiad took place in the Maracanã stadium. While the ceremony sought to celebrate a folkloric idea of the foundations of Brazilian identity, highly grounded in myths constructed by Freyre (2019),[8] that idyllic representation was partially muffled by the crowd booing president Michel Temer – Rousseff's successor – in the stadium.

It was in this political turmoil that the Games were delivered as well as its legacy, whose media coverage will be analysed in the following chapters.

1.3 The Paralympic, Winter and Youth Games and their legacies

Even though this thesis looks at the media coverage of the legacy of the Summer Olympic Games, this is just one of the sporting events organised by the IOC. Since 1924, they have also regularly promoted the Winter Olympics every four years, and the Youth Olympic Games since 2010. Both events are much smaller in regards to the number of athletes participating in comparison with the Summer Games. The Rio 2016 Summer Olympics had 11,238 competitors and Tokyo 2020 had 11,420. In contrast, the Beijing 2022 Winter Games,

---

[8] Gilberto Freyre was a brazilian sociologist-anthropologist. In his book *Casa Grande & Senzala*, published originally in 1930, Freyre broke with the historical tradition by writing about the mixed-race nature of Brazilians in a positive way. His work is, however, highly criticised for creating a myth of a racial democracy (Silva, 2003). Freyre's conceptualisation of the Brazilian people formed as a result of the mixture between indigenous peoples, white Europeans and black Africans is considered to undermine racial inequalities provoked by colonialism in Brazil.

for example, had 2,834 athletes competing, and PyeongChang 2018 had 2,833. The Buenos Aires Youth Games 2018 had the participation of 4,000 athletes.[9]

Although smaller, the cost of these events sometimes far exceeds the cost of the Summer Olympics. The Sochi 2014 Winter Games, for example, cost more than 51 billion dollars, which is almost three times the cost of London 2012 and 2.5 times the cost of Rio 2016. Both the Winter and the Youth Games are also expected to leave a legacy (Chapplet, 2003; Aplin and Lyon, 2014) as they are motivated by the same values of Olympism, promoted by the IOC.

Beyond these two events, which occur independently from the Summer Olympics agenda, there is one that takes place in the same host city, right after the end of the main Games: the Paralympic Games. Organised by the International Paralympic Committee (IPC), it has its origin in competitions organised by British veterans of World War II in 1948. The Paralympic Games have been taking place immediately after the Summer Games since the 1968 Seoul Olympics.

Mentions of the Paralympic Games in this thesis are, however, scarce. Authors such as Dickson et al. (2011), Misener et al. (2013) and Pappous and Brown (2018) have pointed out the lack of scholarly attention to the Paralympic Games as objects of study in the context of legacy. Cashman (2006) argued that the task of distinguishing the legacy of the Olympics and Paralympics can be challenging, because they are not completely independent events. His argument makes sense in regards to the fact that material and immaterial legacies built for one event tend also to be integrated as a legacy of the other. However, from a discursive perspective, a clear separation of both can be identified. The legacy of the Paralympic Games is referred to as 'Paralympic legacy' and tends to be disassociated from the concept of 'Olympic legacy' in the media discourse. Misener (2013, p. 349) looked at the media

---

[9] Dakar 2022 has been postponed to 2026 due to the Covid-19 pandemic.

coverage of the legacy discourses of the 2010 Vancouver Winter Paralympic Games[10]. By employing a critical disability perspective, the author's focus was on the representation of people with disabilities in the newspapers. A critique was made of the 'supercrip' narrative in which Parathletes are framed as heroes who overcome difficulties.

Despite the temporal proximity of both events, Paralympic legacy entails different and more specific attributes, such as upgrades in infrastructure to promote accessibility for people with disabilities in host cities, community awareness-raising about accessibility, and the integration of people with disabilities in the event as volunteers or employees (Misener et al., 2013, p. 334). Therefore, in order to include Paralympic legacy in this research, it would have been necessary to expand the corpus to include specific content on the Paralympic Games.

1.4 Media and legacy: previous work

Works on the media coverage of the legacy of the Olympic Games are scarce. This section introduces some of the methodologies proposed by scholars to analyse news articles on the topic of the legacy of mega-events and their key findings.

Poffo (2012, p. 11) used content analysis to study the news coverage made by the Brazilian newspaper Folha de São Paulo on the legacy of the 2007 Rio Pan American Games. A keyword search was performed on the digitised version of the newspaper to select the final corpus composed of 61 relevant articles. The articles were organised in seven categories based on their framing: urban infrastructure, sporting infrastructure, economy, nationalism, culture and others. For each category, the main legacies discussed in the newspaper were identified. For a second analysis, the articles were categorised by the author according to

---

[10] Four Canadian newspapers – two local (The Vancouver Sun The Province) and two national (The Globe and Mail and The National Post) – were included in this study. From these sources, the author managed to select 62 articles covering the Paralympic Winter Games by using a search engine to explore the digitised version of the newspapers and analysed them using a content analysis approach.

whether they were supporting the legacy of the event, or presenting a neutral or critical perspective.

While Poffo made use of qualitative approaches to study media content, Hedenborg (2015, p. 796) analysed how the Swedish media covered the 2012 London Olympics quantitatively. 1881 articles published online on the website of the newspaper Dagens Nyheter and the tabloid paper Aftonbladet were collected. The quantitative approach focused on the representation of gender and nationality in the media. Articles were categorised based on the gender and nationality of the mentioned athletes. The analysis concluded that there had been an increase in Swedish media attention given to female athletes in comparison with previous Games, while the coverage had centred primarily on Swedish athletes, paying less attention to the other nationalities.

O'Neil and Mulready (2015) also focused on the representation of women's sports in the media. The authors looked at the coverage of the UK national press of women's sports during the 2012 Olympics and one year later. The comparative study sought to analyse whether there had been a legacy of the Olympics with regards to the media attention to women's sports. In conclusion, the change was considered 'minimal' by the authors. This research corpus was composed of seven UK national newspapers: Daily Telegraph, The Guardian, Times, The Daily Mail, Daily Express, The Sun and The Daily Mirror. 4,576 articles were collected from the sports section of the above-mentioned newspapers and only 5% focused on women's sports.

Beyond the focus on aspects of material legacy, Fernandes da Silva et al. (2018) studied 199 news articles published by the Brazilian newspaper O Tempo to investigate the discourses on immaterial legacy. The authors' understanding of immaterial legacy is, however, specific to historical legacy and its potential for corroborating the construction of a national identity. For the analysis, articles were categorised as follows: those covering the history of the Olympics, others on the history of Olympic sports and the last on the life stories of athletes. The paper

does not provide essential information on the methodology such as methods for data collection and analytical tools used.

A focus on positive and negative discourses in the media, such as those analysed by Poffo (2012), was also present in the work of De Hollanda and Medeiros (2020). The authors analysed the coverage of the 2016 Rio Olympics made by the following Brazilian news outlets: UOL, IG, El País, EBC, Folha de São Paulo, Globo.com, Jornal do Brasil, Carta Capital, BBC Brasil, IstoÉ and Época. The authors collected 84 news articles and classified them manually – qualitative approach – as positive, negative or neutral, according to the content narrated by each of the articles. They also conducted a diachronic analysis by dividing the news articles based on whether they were published before, during or after the Olympics. In the same study, the authors analysed opinion surveys to compare with the sentiment of the media coverage. The paper concluded that the articles were more negative before the event and more positive after. This conclusion is at odds with the findings of this thesis, as will be presented in the following chapters. One of the reasons is that the study of De Hollanda and Medeiros (2020) did not only consider articles covering the legacy of the Olympics. Instead, they collected articles covering the 2016 Olympics in general. This study is still relevant to be mentioned here as it shows interest in the sentiment of news articles covering the event and this thesis will look at these sentiments across a larger corpus, using automated methods.

Haw (2022) conducted a quantitative study on the prominence of the term legacy in the British media coverage of the 2012 London Olympic Games, between 2005 and 2017. The corpus was composed of four newspaper outlets: The Daily Mail, The Guardian, The Telegraph and The Sun. 3,000 articles were collected using the database Nexis. The author read each of the articles and categorised them manually as published before, during or after the Olympics as well as based on whether they covered one of the following legacies: socio-cultural, environmental, economic, urban or sporting. The analysis compared the distribution of articles covering the Olympic legacy of London 2012 over time, pointing to an increase during the year of the Olympics and a decrease after. An increase in the number of

published articles was encountered in 2008, the year of the Beijing Olympics, as the legacy of the Chinese event began to be compared to the upcoming British one.

The author also conducted a discourse analysis, in which he identified negative coverage of the event on the topic of environmental legacy prior to the Games, which disappeared afterwards. According to the author, 'political references made within the socio-political domain reflected the political ideologies of the papers reporting' (Haw, 2022, p. 17). The study however did not elaborate further on this argument and just cited a couple of examples, such as xenophobic discourses against immigrants working in the construction of the Olympic village published by the Daily Mail. The paper does not include references to the news articles mentioned, which makes it difficult to look closer at the studied material. To conclude, the author highlighted how studies of this nature contribute to 'developing a long-term media strategy when bidding to host sport mega events' (Haw, 2022, p. 19). The author continued by stating that 'such a strategy should ensure national media outlets are able to report in an unbiased but sensitive discourse around the hosting of global events' (Haw, 2022, p.19).

As there is no such thing as unbiased journalism (Moraes, 2019), studies like Haw's and my own might instead contribute to more informed decision-making by governments and the IOC on their communications strategy, improving how the legacy plans are presented and officially communicated. These studies can also help in the construction of future legacy plans, by looking at how the media has previously covered the legacy of the Olympics, in order to highlight positive aspects that tend to receive more attention.

1.5 The contribution of this thesis

The contribution of this thesis is twofold: theoretical, regarding the conceptualisation of Olympic legacy based on the media coverage of the London 2012 and Rio 2016 Olympic Games; and methodological, proposing a pipeline for analysing news articles using Natural Language Processing (NLP) techniques combined with a qualitative discourse analysis.

As discussed above, studies on the media coverage of legacy are scarce and have been focused on a limited number of data sources. Most papers published on the topic have used a qualitative content analysis to identify specific aspects of the coverage – such as the coverage of women's sports – or a quantitative analysis to account for the temporal and categorical distribution of news articles published by news outlets. Haw's (2022) was the only study to have used a mixed methods approach. His study combined a quantitative study of the news articles covering the legacy of the 2012 London Olympics with close reading, conducted through the use of discourse analysis methods to investigate some of the topics discussed in the media. Haw's work makes a valuable contribution, especially concerning the laborious work of close reading 3,000 articles. However, for such an effort of close reading, the paper presents limitations such as the lack of a qualitative analysis, which could have provided meaningful insights with regards to the numbers and classifications presented by the author.

None of the works published on the media coverage of Olympic legacy have conducted a comparative study of the coverage of two events. The comparative study allows for the comparison of how different cultural media practices, social, political and economic contexts, and linguistic practices shape the legacy discourse. While enriching, comparative studies of this nature also present methodological challenges such as multilinguality and the discrepancy regarding the availability of data sources, which are discussed in the following chapters.

As pointed out by Thomson, Schlenker and Schulenkorf (2013, p. 119), there has been limited consensus on the concept of legacy with regards to sporting events. This thesis also contributes to the conceptualisation of legacy but from a discursive and mediatic perspective. By looking at how the legacies of two Olympic Games, London 2012 and Rio 2016, were framed by the British and Brazilian media, this study constructs a conceptualisation of the narrated Olympic legacy, which differs from the IOC's definition, as well as from the definitions provided by scholars. Academic work produced in disciplines such as political

science, sociology, philosophy, sports and urban studies[11] has looked at the legacy of sporting events from different perspectives, analysing public opinion, official debates promoted by the IOC, documentation produced by governments on legacy plans and financial expenditure, among others. This work, however, focuses on news media discourses and how they construct a perception of what legacy is, was and will be. It draws on the work of authors such as Vizeu (2009), for whom the media is understood as not solely a mechanism for reporting the constitutive facts of reality but instead as an active actor in the construction of the perceived reality. In this sense, analysing the news media discourse allows us to comprehend what the reference for accessing a frame of social reality is as well as the effects produced on the concept of legacy, based on how information is conveyed.

As a thesis produced under the PhD programme in Digital Humanities, this work is the first to make use of computational tools to study news articles covering the concept of Olympic legacy. It also offers relevant methodological contributions. It presents an innovative pipeline for the use of sentiment analysis, from the selection of the most suitable algorithms to the evaluation of their limitations. It proposes not only an exploratory use of quantitative digital methods to track topics of interest to be further investigated using qualitative approaches such as discourse analysis, but it goes further by making use of activist blogs and official government websites to contrast and identify silences and gaps in the news media coverage.

Therefore, the main research question addressed by this thesis is: what is the legacy of the London 2012 and Rio 2016 Olympic Games as conceptualised through the British and Brazilian media coverage of these events?

The objectives of the thesis are:

- Identifying in the news articles covering the legacy of the two events what aspects of Olympic legacy have received attention;

---

[11] See, for example, Chappelet, 2003; Hall, 2006; Gratton & Preuss, 2008; MacAloon, 2008; Bairner & Molnar, 2010; Dickson, Benson & Blackman, 2011; Leopkey & Parent, 2012; Castro & Novaes, 2014; Borges & Santos, 2018; Preuss, 2019; Girginov, 2018.

- Finding evidence of silences in the news media discourses by comparing them with activist and government discourses;

- Conceptualising Olympic legacy based on the aspects of legacy (material and immaterial, planned and unplanned, positive and negative) covered by the news media;

- Providing researchers in the Digital Humanities with a methodological pipeline to collect and analyse multilingual news articles on a specific event by combining quantitative and qualitative approaches. Special attention is given to how quantitative and qualitative methods inform each other and enrich potential conclusions.

As an author, I also have personal interests in this topic. Even during my undergraduate studies in Journalism, I demonstrated an interest in studying digital media, starting with social media analysis. In 2013 and 2014, Recife, the city where I lived, was the stage of protests against the decision-making power of private capital in the use of urban public spaces. My interest in debates on urban rights and online activism led me to a Master's degree in Communication, where I studied the communicational organisation of activists of a local Occupy movement. Besides my academic involvement with the topics of digital media and urban studies, I have also been directly involved in two of the major sporting events hosted in Brazil in recent years. I worked in broadcasting production for the 2014 FIFA World Cup and the 2016 Rio Olympic Games.

When I joined the CLEOPATRA project to work on the vast amount of content produced about the London and Rio Olympics, legacy rapidly became my main topic of interest due to its political implications. I was interested in the impact of hosting major sporting events and the debates around the promises of improving urban infrastructure and social services in the host cities. This work is, therefore, motivated by my personal experience of having witnessed both the magical atmosphere of the Games in Rio and the struggles of Brazilian citizens due to exclusionary policies of urban regeneration.

1.6 Overview of the chapters

This thesis is divided into eight chapters.

In this Introduction (Chapter 1), I provided a contextualisation of the research topic and a brief background history of the Olympics, with particular focus on the London 2012 and Rio 2016 Games.

In Conceptualising Olympic legacy (Chapter 2), a literature review of the topic of Olympic legacy is provided. This chapter discusses how different authors and entities have conceptualised the dimensions of legacy. I cover the debates on what a media event is and how the Olympics can be understood as one. The chapter presents a reflection on the role of the media as a place of reference and how it contributes to the development of an idea of what legacy is and what it is not. I also offer a conceptualisation of the emotions of the legacy discourses and how utopian and dystopian narratives are constructed by the media. To conclude, the chapter discusses the temporality of the concept of legacy, and the challenges of defining when it begins and when it ends.

The Methodology (Chapter 3) starts by presenting the research pipeline design and the construction of the dataset, including how decisions on the relevant data sources to be used and the processes of data collection were taken. A reflection on the limitations of accessing the data and the data itself is also presented. Next, a sub-curated dataset is produced for the qualitative analysis and the processes of constructing this corpus are provided. The chapter presents a detailed explanation of the NLP tools used for this study, including sentiment analysis, as well as the method of network analysis. To conclude, I explain how this work has used the method of discourse analysis to inform the qualitative close reading performed on the data.

In The British media narrative of London's legacy (Chapter 4), the results of the experiments performed with the British News Media Corpus, composed of news articles published by the British media on the legacy of the 2012 London Olympics, are discussed. Aspects of the media coverage are explored using NLP, network analysis and discourse analysis. Among the

main aspects discussed are the disputes over the use of the Olympic stadium; the critical tone of the coverage before the event; the news media response to the legacy plans, especially with regards to the projects for young people; the cyclical nature of the legacy coverage, in which journalists make use of past examples to project possible futures; and the role of important political figures such as Boris Johnson in the legacy discourse. In this chapter, I also discuss the results of the analysis of the British Government Corpus to investigate the government's framing of legacy in texts published on their official websites; and of the British Activist Corpus, to study the media silences and the activists' warnings regarding the negative impacts of hosting the event. The analysis covers the conflicts between a project of inspiring people around the globe versus the media interest in local issues. A discussion of the news outlets' economic organisation and the characteristics of their coverage is also provided.

The Brazilian media narrative of Rio's legacy (Chapter 5) presents the results and discussion of the experiments performed on the Brazilian News Media Corpus, composed of news articles published by Brazilian media outlets on the legacy of the Rio 2016 Olympic Games. It begins with a reflection on the methodological challenges of finding and collecting the data from Brazilian sources and the limitations of NLP tools for working with the Portuguese language. Then, it analyses the positive discourse of Rio authorities regarding their expectations of transforming the city through the Games; the predominance of words related to the justice system in the corpus and the influence of the Brazilian political crisis on the media coverage of legacy; the media focus on the challenges faced by Rio to be prepared for hosting the Olympics, mainly with regards to the decontamination of bays and building infrastructure; the media coverage of the ruins of legacy, especially regarding venue fires and the political debates about solving the issues by privatising the infrastructure; and the political trajectory of Eduardo Paes, Rio's mayor, and how it has influenced the legacy discourse. Based on the Brazilian Government Corpus, I investigate how the Brazilian and Rio governments disputed the legacy narratives; and, by referring to the Brazilian Activist Corpus, I reveal the response of activist blogs to gentrification and their interaction with global discussions on the impact of hosting the Olympics. Chapter 5 also analyses the Brazilian media outlets and how their content differs from one to the other. At the end, I

suggest the concept of a 'gambiarra' Games, as a way to understand how the event has been narrated.

In Language, sentiment and legacy: an affective dispute (Chapter 6), the result of the experiments applying sentiment analysis to the news articles is presented and discussed. The chapter debates how useful analysing sentiment in news articles can be and what it potentially reveals about the media coverage of certain events. I also discuss how sentiment has changed over time in the British News Media Corpus and in the Brazilian News Media Corpus, and how this change may be interpreted as the 'scepticism of Londoners and the optimism of Brazilians'. The way desire is manifested in the legacy discourse is critically approached and, to conclude, I argue that sentiment is expressed in the articles as part of an affective dispute.

Chapter 7, Towards a conceptualisation of the narrated Olympic legacy, presents the key aspects of the legacy concept based on how it has been approached by the media. I revisit Gratton and Preuss's (2008, p.1924) definition of Olympic legacy, offering a new conceptualisation. This chapter also reflects on the use of news articles as objects of study, focusing on their unstable and complex characteristics.

The concluding Chapter 8 provides a summary of the key findings, a reflection on the methodology, and an overview of the practical implications of this work, its limitations and suggestions for future work.

To conclude this introduction, I would like to briefly comment on some stylistic choices that were made necessary to improve readability and accessibility of the content of this thesis:

1) News outlets: when news articles are cited, they will include the name of the author, the name of the news outlet – publisher, and the year in which the article was published as follows: (Author, Outlet, Year). When the news article does not include the name of a journalist, it will include only outlet and date of publication as follows: (Outlet, Year). The decision to include the outlet was to make it easier for the reader

to identify some of the particular characteristics of the news publishers without necessarily having to check the list of references at the end of the thesis while reading.

2) Archived URLs: all websites cited in this thesis were archived by the author or referred to an already archived version in the Internet Archive. During the four years of study, many of the news articles collected for analysis were deleted from the live web. This method allowed for the preservation of the material and for a long-lasting list of references. In the list of references, you will find the original URL of the news articles accompanied by a link for the archived version.

3) Charts created by the author for this thesis have been designed to be accessible in colour terms, apart from those images that result from screenshots taken from platforms such as TIME (a CLEOPATRA project demonstrator).

Chapter 2: Conceptualising Olympic legacy

As soon as a city launches its candidature to host the Olympic Games, the question that emerges at the centre of the bidding process is: what are the benefits of hosting one of the biggest sporting events in the world? Usually, the concept of legacy is invoked in order to legitimise the decision to participate in this very controversial process.

Despite being a pillar of the Olympic Movement, and having been regularly brought up to justify cities' – and nations' – participation in the event, the concept of legacy is not very clearly defined and it continues to require effort by scholars and members of the IOC to determine – or even get closer to determining – exactly what it means (Preuss, 2007; Cashman, 2005). Gratton and Preuss (2008) described three reasons why a positive legacy is one of the main concerns of the IOC: it prevents people from blaming the IOC after the event, it justifies the use of scarce public resources and it also encourages other cities to bid for future events. This last reason is fundamental to guaranteeing the future of the Olympic Games.

Recently, many cities have given up participating. In an article[12] published on 6 February 2018 in BusinessInsider.com, Rob Ludacer came up with some reasons why we might be watching what he calls the 'death of the Olympic Games'. First, it is worth noting how the number of bids has reduced over time. While for the 2004 Olympics, there were 11 cities competing to host the event, in the recent 2022 and 2024 bids, there were only two[13] and three[14] respectively, because most of the initial candidates withdrew from the bidding process.

---

[12] Available at
<https://www.businessinsider.com/future-olympics-no-country-wants-to-host-games-2018-2?r=US&IR=T>. Accessed on 2nd June 2020.

[13] Beijing (China) and Almaty (Kazakhstan).

[14] Paris (France) and Los Angeles (USA). Hamburg (Germany), Rome (Italy), Budapest (Hungary) and Boston (USA) withdrew their bids.

Only Beijing and Almaty applied to host the 2022 Winter Olympics. Second, the article mentions the huge difference between the budgeted and final costs. For the Sochi 2014 winter games, the planned investment of US$10.3 billion was exceeded, and it ended up costing approximately US$51 billion. For London and Rio, the difference between planning and execution in terms of money was not that large, in comparison with the previous Russian example. For London it went from US$5 billion to US$18 billion and for Rio from US$14 billion to US$20 billion (Ludacer, 2008). The broader context is even worse, considering that Chicago spent almost US$100 million on the campaign to host the 2016 Olympics and ended up losing the bid[15].

Another aspect to be taken into consideration is the change in the profit that the host cities make from TV licensing rights. Ludacer (2008) cited as an example that during the 1990s the IOC took 4% of revenue, and the rest went to the host city. For the Rio 2016 Olympics, the IOC cut rose to 70%. Finally, something that seems to be threatening the future of the Olympics, at least the Winter events, is climate change. Most of the cities that hosted these events in the past no longer have enough snow to put on a successful games.

Facing this challenging environment, Olympic legacy remains crucial for the continuation of the games. It is part of a 'legitimating narrative' (Poynter and MacRury, 2012, p. 315), especially because, as mentioned above, it justifies the high costs of the event. The idea here is that this is not 'just' a sporting event: it is rather a platform for development. The term legacy has become a standard feature of Olympic discourse as it represents a balance between the promise of a Games – what is expected – and the municipal economic realities – what is delivered (Poynter and MacRury, 2012).

Reinforcing the multiple meanings the word legacy encompasses as identified by Cashman (2005), MacAloon (2008) has pointed out that the concept of legacy became a kind of

---

[15] Although the Olympics have been facing a huge crisis, there are a few cases of Olympic legacies that are considered models of success in modern Olympic history, such as Sydney 2000 or the most recognizable example, Barcelona 1992. It is surprising, though, that the word legacy appeared only three times in official reports of the Barcelona games (Girginov, 2018).

magical discourse in IOC public relations once it assumed a 'cross-functional, cross-contextual and transnational hegemony' (MacAloon, 2008, p. 2061). The first time the word legacy was recorded in the context of the Olympics was in the documents of the 1956 Melbourne Olympics bid (Girginov, 2018). Besides claiming its transnational reach, MacAloon (2008) also highlighted a slight – but significant – variation in the concept of legacy across language borders. The IOC has two official languages, English and French. French speakers use the term *héritage* rather than a direct equivalent to the English 'legacy'. Both words carry a similar meaning but according to the author, while in *héritage* we have a sense of something that is transmitted from the past to the present, 'legacy' privileges what is being planned for the future. In Vancouver 2010, the slogan 'legacies now' was an attempt to subvert this logic of delivering in the future by proposing to offer benefits during the entire project. Inspired by Girginov (2018), this work will look at legacy as a prospective concept – rather than a retrospective one. The main point here is to understand Olympic legacy as a process that is not concluded when the Olympics as an event ends, but remains over the years as a reference for urban development in the host cities.

Looking back at the 2008 and 2012 Olympics, there are two different motivations for staging the games that stand out and expand understanding of the concept of legacy. For London at that moment, hosting the summer Olympics represented the chance to remake the city for the future. For Beijing, on the other hand, the event was an opportunity to communicate years of economic development that had already been achieved (Poynter and MacRury, 2012). These two different perspectives highlight the temporal haziness in which the concept of legacy stands.

2.1 The dimensions of legacy

By analysing the word legacy etymologically – it means 'property left by will, a gift by will' – the sense of a present being given is easily identifiable. This might explain why official narratives of governments and organising committees have historically been conducted in a beneficial or even 'generous' tone. For Cashman (2005), the word legacy is problematic and

dangerous (p. 15): first, because it is assumed by the organising committees to be necessarily positive; second, because it carries the sense of something that is naturally transferred to the community at the end of the games; and finally, because legacy is taken as word that does not need to be explained – it is self-evident. The main goal of this thesis is to offer a conceptualization of the word legacy based on its use in the Brazilian and British media coverage of the London 2012 and Rio 2016 Olympics.

Based on the studies of Gratton and Preuss (2008) it is possible to identify the existence of at least three aspects of legacy. First, there are tangible and intangible legacies. Among the most recognizable legacies of the Olympic Games are sports infrastructure – such as new stadiums and training venues – and urban planning – which involves many aspects, such as new residential areas and new transportation infrastructure. But the concept is not reducible to material – or hard – legacy. There are also many abstract/immaterial legacies that might be described as soft legacies. As an example one might think of national self-confidence (MacAloon, 2008, p. 2064), the renewal of community spirit, the improvement of interregional cooperation, the production of new ideas, popular memory and additional know-how (Gratton and Preuss, 2008, p. 1923). Second, it is important to point out that legacy does not have to be positive, although most of the time it is used in positive contexts by governments and organising committees. Gratton and Preuss (2008) have called attention to possible negative legacies such as 'debts from construction, high opportunity costs, infrastructure that is not needed after the event, property rental increases' and so on (p. 1923). For that reason, some authors prefer to use the word 'outcomes' rather than legacy, as an attempt to achieve a more neutral discourse (Preuss and Holger, 2006). Third, there is another dimension of legacy, which is the existence of unplanned legacies that occur as a consequence of other events. Based on these three 'properties' of legacy, the authors offered a definition in which they argue that 'legacy is planned and unplanned, positive and negative, intangible and tangible structures created through a sport event that remain after the event' (Gratton and Preuss, 2008, p. 1924). These categories will be referred to later on in this work as a starting point for data analysis.

These multiple dimensions of the legacy of sporting events are evidenced by their political appropriation. In December 2019, the UK Prime Minister, Boris Johnson, tweeted his desire to stage the 2030 FIFA World Cup. 'I want it to show our national confidence as we get Brexit done', wrote Johnson. National self-confidence was mentioned above as an example of immaterial legacy, but this event also highlights the political dimension of legacy, since sporting events 'constitute a key component of domestic and international development strategies' (Nauright, 2004, p. 1325).

As pointed out by Tomlinson and Young (2006), studying sporting spectacle as a media event cannot be dissociated from the study of power. As a form of cultural history, its narrative is permeated by different 'conceptions, values and ideologies' (Tomlinson and Young, 2006, p.4). Although the political use of the Olympics is a commonplace in history, it was only in the 1990s that being a host became 'explicitly linked with a range of social, economic and sporting benefits as a more tangible measure of the three forms of cultural capital' (Girginov, 2018, p.3). This will be evidenced through further analysis of the appearance of political entities, politicians and government bodies in different news articles, demonstrating how Olympic legacy might be considered the most significant expression of the political and economic dimensions of global sport.

2.2 Olympic legacy over time: delimiting a media event

In this section, I aim to contextualise the Olympics, and their legacies, as media events and reflect on the consequences of this mediatisation. The political and economic dimensions, discussed above, are mostly of interest to researchers in sociology, sports history, economics, and urbanism. These scholars have analysed official documents published by governments, the IOC, and other bodies, to investigate how legacies came about. There is also, however, the discursive dimension of legacy, which refers to how this concept is textually constructed.

The Cambridge Dictionary describes the word 'event' as 'anything that happens, especially something important or unusual'. The dictionary provides three examples: '(1) A planned

public or social occasion (2) Each of several particular contests making up a sports competition (3) Physics: a single occurrence of a process, e.g. the ionisation of one atom'. What social occasions, sports competitions and single occurrences of a process in physics have in common is the delimited space in time during which they occur. When it comes to media events, by contrast, the boundaries of time may not be very well defined.

According to Dayan and Katz (1992), a media event is recognized by the audience as a disruption in their daily routines. The authors argued that media events can differ from other types of broadcasting because their interruption is monopolistic and they happen live – and are therefore unpredictable: 'Figuratively, at least, these events induce people to dress up, rather than dress down, to view television' (Dayan and Katz, 1992, p. 9). Although the authors offered a seminal contribution to research on media events, their work, written in 1992, has many temporal limitations. However, analysing these limitations helps us to reflect on how media events have been affected by the development of new and more sophisticated technologies, from advanced broadcasting equipment to the emergence of the World Wide Web. Faster technologies have accelerated the information flow and today it is very difficult to think of an event as something that is necessarily 'preplanned, announced and advertised in advance' (Dayan and Katz, 1992, p. 7). Broadcasters no longer need much time to plan live coverage of an event. Only a few minutes after the beginning of the Notre-Dame fire on 15th of April 2019, it was possible to watch a live video transmission of the event around the world.

This thesis is dedicated to the study of an event which is extremely planned, but looking at the Olympics, and more specifically at their legacy, the temporal boundaries are not precise. London 2012 and Rio 2016 will be considered here as two different events – given their socio-political-economic-territorial particularities – that start with the candidature to stage the games and have an unclear conclusion, considering that the impact of both is still reverberating in the media.

The Olympics have been taken as a standard example for authors studying media events. Dayan and Katz (1992) used the Olympiad as an example of one of the three basic scripts for

media events: 'contest'. Among the main characteristics of this script are the cyclical order in which events occur; the agreed rules under which the event takes place; the drama of 'who will win?'; a certain neutrality that gives the audience the role of judging; and the focus on the present. One of Dayan and Katz's students, Eric Rothenbuhler (1985) took the 1984 Olympics as a case-study to investigate whether it 'qualified as a celebration of a coherent set of values, beliefs and symbols' (Hepp and Couldry, 2010) by bringing together the public around a media event. What is interesting to note is that the conclusion of these authors was that there was an inconsistency between the games and their values in the way it was perceived by American society. Rothenbuhler's study raised an important critique of Dayan and Katz's idea that the ritualistic characteristics of media events promote social integration around a common set of values.

Hepp and Couldry (2010) updated the concept of media events, reframing the discussion within the globalised world. They divided the critiques of previous approaches into three points. First, there is the ritual perspective: unlike Dayan and Katz's idea of a homogeneous integration of values, Hepp and Couldry took the Olympics as an example of an event which, depending on the region and nation, will be mediated in a very different way. This perspective deconstructs the 'myth of the mediated centre' (Hepp and Couldry, 2010, p.9) as a hegemonic discourse. Second, they considered media events as genres: the authors raised a critique of the way media events were considered as happening independently of real events, an idea that is not sustainable from the perspective of mediatized cultures. Therefore, they do not refer to 'discourses about an event' but rather Firske's (1994) conceptualization of 'discursive events' (Hepp and Couldry, 2010, p. 11). Third, they identified the narrow typification of media events: Hepp and Couldry extended the understanding of media events by highlighting the influence of celebrity cultures in a form of eventization. This is the way reality shows and tabloids create an event around the life of a celebrity, which goes beyond the definition of rituals.

These multiple readings of the concept of media events have shown the importance of the media not only as merely mediums for the reproduction of such events but actually as active

producers of discourses. In a mediatized society, events are discursive as they are permeated by a diverse range of social-cultural-political and territorial narratives. More than a disruption in the everyday routine where individuals would unite to 'watch' a single event, media events are susceptible to counter-discourses, reinterpretations and reframing of their official narrative: they are a space of dispute.

2.3 Legacy as a media narrative

Bairner and Molnar have written about the different uses and framing of the word legacy as the 'imagineering of legacies', the idea of a collective mediated interpretation of the phenomena: 'The media help to "imagineer" images of host locations as "world class" - for people inside as well as outside the host city and country' (2010, p. 37).

Thinking of legacy as media narrative is a reflection on the role of the media in the construction of legacy storytelling. This thesis looks at news articles as key elements in the way the concept of legacy is perceived and interpreted. As pointed out by Couldry and Hepp (2017), a reflection on the mediated construction of reality can be problematic in many ways, due to the historical theoretical changes that have affected the way scholars understand both the term media and the term reality. This thesis also considers Latour's (2012) critique of social theory to develop the idea of a mediated constructed legacy that takes into account the role of not only news producers but also news 'consumers' and non-human agents - technology - as fundamental actors in the construction of meanings through multiple processes of interpretation.

There are two important points raised by theories of journalism in terms of the way journalistic practices impact the coverage of an event and how we interpret those discourses: the myth of objectivity in journalism and the concept of journalism as a place of reference. Poerksen (2008) outlined some of the contributions of constructivist epistemology to the understanding of journalistic subjectivity through the philosophy of language. First, the author pointed out the selection of information as the main activity of journalism as it is 'an

act of active construction, a form of giving meaning, a form of guided interpretation' (Poerksen, 2008, p.297). This selection is followed by decisions about the words used to narrate the story and also the way in which media genres, language patterns and forms of presenting the information shape the communication process. It is important to mention here the contributions of authors such as Michel Foucault (1975) and Pierre Bourdieu (1992) to the understanding of language as pervaded by ideology and, therefore, power. The work of these authors highlights the inevitable subjectivism of everyday communication. From the constructivist perspective, objectivity would then be a much more unrealistic and unreachable ideal to be taken as a ritual for journalists than the supposed neutrality evoked by media outlets. This comprehension of the role that language plays in journalism is relevant for this research in reflecting on the texts not merely as descriptions of reality but as actors in the active construction of this reality and, therefore, of the way the history of these events will be narrated.

In the appendix of his book *On Television*, Pierre Bourdieu (1998) highlighted the discursive aspect of the Olympic Games, as he argued that the event is produced twice: 'The first production is the actual event in the stadium (...); the second show reproduces the first in images and commentary' (Bourdieu, 1998, p.81). Bourdieu's perspective on discourse analysis is fundamentally based on sociolinguistics. According to Van Dijk (2008, p. 192), sociology has profoundly impacted the study of discourses and one of its biggest contributions is the concept of ideology as discursive. This approach to discourse analysis considers that ideas and beliefs are produced and reproduced through communication. Van Dijk (2008, p. 193) understood ideology as 'axiomatic beliefs', and therefore fundamental values shared by a social group. This reflection on ideology and discourses frames the way news is comprehended as it is disassociated from the positivist objectivism – highly endorsed by the journalistic community – to be seen as a cultural product profoundly affected by authorship. Myles (2010) in his book *Bourdieu, Language and the Media* analysed the contributions of *Language and Symbolic Power* (Bourdieu and Thompson, 1991) to the study of media discourses. The understanding of the media as an interested party and, therefore,

operator of symbolic power is key to any analysis of the journalistic narratives in the coverage of events.

Expanding on Poerksen's constructivist perspective, Vizeu (2009, p. 77) brought out the concept of 'natural attitude of the life-world' postulated by Alfred Schütz to develop the idea of journalism as a 'place of reference'. For Vizeu, journalism plays an important role in society as it reinforces the existence of reality and also that we are not alone in the world. This is what Schütz has called 'the practical taken for granted world' (1962). As part of an everyday life routine, news consumption can be understood, according to the author, as a 'site' in a very similar way to religion, schools, family and friends. It is a 'site' where people go in search of stability. Vizeu's theory of 'journalism as a site of reference' contributes to the way the role of journalism in our society is understood. While his work is mainly focused on television, it provides a theoretical approach that can be used to reflect on the role that websites like Google play in our society nowadays[16].

Returning to the concept of narrative, Roland Barthes, in his 'Introduction to the structural analysis of narrative', stated that 'no one can produce narrative without referring himself to an implicit system of units and rules' (1975, p. 238). On similar lines, White described narrative as a 'metacode, a human universal on the basis of which transcultural messages about the nature of a shared reality can be transmitted' (White, 1980, p. 6). Thinking of legacy as a narrative is precisely referring to this unstable and momentary set of signs, shared by individuals, that feeds journalists and is reconstructed by them through the act of writing stories.

---

[16] Google is the second most accessed website for news consumption in the UK (Ofcom, 2022, p.52). Given its importance, Google search results were used as a key reference for collecting news on the legacy of the Olympic Games for this study.

2.4 The boundary between a utopian and dystopian legacy

The article entitled 'The Olympic Games is a fantastic excuse to change Rio' (Globo, 2012b) published by the Brazilian online news website G1.com reflects the words uttered by the mayor of Rio, Eduardo Paes. For Paes, it was the weakness of Rio's infrastructures that made the city the best candidate for hosting the games: 'We told the IOC that Chicago, Madrid and Tokyo are fantastic cities (…) but if the Olympics is about transforming a city, building a legacy, changing people's lives, then it must happen in Rio'.

Legacy as a concept is often used to express the desire of a utopian city to be achieved by using the Games as a development catalyst. As stated by Girginov, the concept is a social construction that promotes 'specific visions of what the desired development should look like' (2018, p.196). This idea of desiring a city was deeply addressed by David Harvey in his writings on urban activism. As claimed by Harvey, 'city' carries a historical symbolism that attributes political layers to the term. The author described the city as 'an object of utopian desire, as a distinctive place of belonging' (Harvey, 2012, p. 16). This utopian thought has been crucial in urban planning and has contributed significantly to the transformation of urban landscapes worldwide (Pinder, 2002).

Utopian thought has also been important to the legacy discourse. For Girginov (2018) legacy follows the three-tasks stated by Snow and Benfored (1988): diagnostic, prognostic and motivational framing. This means that in order to provide a legacy plan, issues need to be identified, solutions must be designed and finally, actions must be performed. It is in the last of the three tasks that utopian thought plays an important role as it is potentially motivational. As pointed out by Snow et. al. (2018), motivational framing for social movement theory is what makes the action to overcome fear. In the context of Olympism, the potential for urban transformation of the legacy is one of the main arguments that supports the Olympic Movement. According to the IOC's website, the main objective of the Olympic Movement is 'to contribute to building a peaceful and better world by educating youth through sport practised in accordance with Olympism and its values' (IOC, 2020). The Olympic Movement is first and foremost a set of ideals that aim to inspire, motivate and guide the actions of

entities committed to delivering the games. As legacy, the Olympic Movement embeds the utopian narrative of transformation. However, for Girginov (2018) the practical use of the concept of legacy does not contribute very effectively to the Olympic Movement, due to its contradictory characteristics, which will be addressed in the following chapters.

Reinforcing Harvey's statement, Pinder (2002) pointed out that cities are constituted by dreams and desires. But the author also includes fears, which will manifest in the expectations of disastrous legacies. Like the concept of city, legacy is also under dispute and in constant tension (Girginov, 2018, p.195). In January 2011, more than one year before London's Olympics, the Dailymail.co.uk published an article entitled 'Mike Lee warns Olympic Stadium will become "white elephant" if Spurs' proposal is rejected'. White elephant is a common expression found in journalistic texts that report negative legacies of major sporting events. It usually refers to those infrastructures constructed for the games that were abandoned right after the event or remain in a precarious condition, not serving the purpose for which they were initially planned. For Leopkey and Parent (2012), white elephants are among the most representative negative aspects of the Olympic legacies. Authors like Papanikolaou (2013) have even called Athens the 'White Elephants Olympic City'. In order to minimise the chances of ending up producing these white elephants as a consequence of the Olympics, the IOC initiated a conference in 2002 to discuss legacy planning (Gratton and Preuss, 2008). Since white elephants have become frequent in Olympic history, it seems to now be part of the imaginary of how the Olympic legacy will look years after the event. It is a constant threat surrounding those responsible for building the games.

In April 2016, months before Rio's Olympics, the *Estadao.com.br* published an article describing the constructions for the event as 'delayed and of dubious quality'. All this fear brought to the fore by the media articles rivals the optimistic utopian discourse of a positive legacy. They highlight those pessimistic aspects that project the legacy discourse towards a much more dystopian scenario. This dystopian idea became prominent in expressions used by authors like Zimbalist (2017), who initiated his book's introduction about the Rio Olympics with the sentence 'welcome to hell'.

It is important to take into account the spatial-temporal dimensions in which these utopian or dystopian projections occur. Spatially, it seems like the boundary is not very precise, as optimistic and pessimistic articles share space many times in the same media outlets. A more profound analysis on this point will be conducted in the following chapters. However, the temporal aspect looks much blurrier. If it is more understandable that before the Olympics, projections have more space to be expressed as they refer to the future, it becomes challenging to understand how these projections relate to the present. Once the games are concluded and the legacies of the events are 'concrete', it is important to reflect on how the reports about the outcomes vary in the way the media describes or criticises them. This research will look more closely at how optimistic and pessimistic scenarios develop over time, in order to understand how they change before and after the event.

To conclude, this thesis investigates who the actors that perform utopian and dystopian discourses in the media are. How different are the perspectives of a politician, an activist and a political commentator about the impact of hosting the Olympics and what is the space designated to each of these dissonant voices in the media outlets' pages? Moreover, it is important to understand what role the location of the event plays in the choice of terms selected to report the games due to the significant socio-political differences between the two cities studied in this project.

As pointed out by Papacharissi, affect is subjectively experienced emotion that plays an important role in promoting engagement. 'Affect contains participation, promise, hope' (2016, p. 6), elements that are intrinsic to the Olympic legacy discourse. Reflecting on the utopian discourses mentioned above, it seems clear that affect is important for sustaining this pillar of the games, aiming to engage multiple actors in the political arena but especially the audience and the citizens of the host city.

2.5 The legacy timeline

Returning to the previous discussion about the different interpretations under which the word legacy falls – due to the two official languages of the IOC –, it is possible to state that the concept requires a closer look in relation to its temporality. The discrepancy between legacy and heritage is not merely linguistic but also ontological, as it reflects different 'worldviews about the nature of reality and legacy in particular' (Girginov, 2018, p.198). For Preuss (2015), legacy is necessarily a result of the event, therefore it must occur after the games. This understanding would be closer to the meaning of the word in English. The author called the 'legacy' achieved before the event, such as structures constructed a long time before the event is staged, merely 'pregnancy effects'. Girginov (2018) in contrast considered everything that is constructed for the games as part of a developmental plan that is responsible for determining - intentionally - what the legacy visions to be considered as relevant are.

Besides inconsistent interpretations of the concept in terms of what time it refers to, temporality also influences the social, political and economic context in which legacy emerges. There is usually a long temporal gap – 7 years – between the election of the host city and the staging of the games. This is enough time to completely change the political situation of the host cities. As pointed out by Girginov (2013), while London's bid was prepared by a Labour government, the 2012 Olympics happened during a Coalition government, which according to the author, provoked significant changes in the legacy delivered. In Brazil, the change was even more dramatic. When the famous scene of President Lula (Workers Party) celebrating the winning of the bid in Copenhagen 2009 was broadcast, the country had achieved a considerable increase in its economic growth. In 2016, his successor, President Rousseff, was impeached, Brazil fell into a huge economic crisis and months after the games, the governor of Rio, responsible for preparing the bid with Lula, was arrested because of corruption.

The Olympics usually occur in very special conditions, when governments tend to be very flexible about many aspects such as economic spending. This does not persist after the event

though (Girginov, 2018). Legacy is then mostly delivered before the end of the games. After that, the restricted resources have a huge impact on the continuation of legacy development (Girginov, 2018, p. 200). To counter this impression, in November 2019 the Brazilian government founded the Office for Olympic Legacy Governance, even though many of the entities responsible for building up and managing the Olympic spaces had been dismantled in Rio at the end of the event. The new office was created to replace the previous AGLO (Olympic Legacy Governance Authority), that had closed in June 2019, as an attempt to communicate that Rio's legacy plan was not abandoned (Globo, 2019b).

To add more complexity to the concept, legacy must also be read through people's interpretations: how people see the effects of a legacy. For Preuss (2015), sometimes it may be felt only a long time after the event. For the author it depends on how the opportunities that emerged from the event are transformed into real benefits. The title of the legacy planning report for London 2012, published by the Department for Digital, Culture, Media and Sport (DCMS) - 'Before, during and after' - reframes the temporal conception of legacy. As pointed out by Poynter and MacRury (2009), this broad understanding that takes into account the past, the present and the future highlights a narrative conception that no longer considers the games as ends in themselves.

A new phenomenon has recently been added to the temporal perspective. Tokyo 2020 was postponed to 2021 because of the Covid-19 pandemic. The break in the routine of 4 years between summer games drew attention to the negative impacts of hosting the event. If it had not previously been easy to sustain the argument of a range of benefits for the host city, the postponement of Tokyo has heightened the rhetoric of activists such as Jules Boykoff, as pointed out by the Washington Post in July 2020 (Donahue, 2020),[17] who said that the Olympics lead to 'overspending, militarization of police, citizen displacement, greenwashing

---

[17] Available at
<https://web.archive.org/web/20230601161530/https://www.washingtonpost.com/magazine/2020/07/06/inside-troubling-legacy-displacing-poor-communities-olympic-games-one-villages-resistance-brazil />. Accessed on 20th October 2020. This page is not available on the live web anymore (checked on the 14th of October 2023). It was reposted by the Rio On Watch website, which is included in the bibliographic reference.

and corruption'. In the recent history of the modern Olympics, the event has rarely broken the 4 years routine.[18] On the one hand, the costs to delay the games are estimated at $2bn-$6bn, according to the Guardian report of July 2020 (McCurry, 2020). On the other hand, Tokyo's 2020 president, Yoshiro Mori, said the cancellation of the event would have cost double its actual cost. The Guardian's report also brought up the result of a poll in which 51% of Japanese people surveyed thought that the postponed event should not be held in the following year – 2021. As argued above, the legacy discourse also depends on the way the public sees and interprets the event's legacy. It is beyond the scope of this thesis, but, in any future investigation special attention will need to be paid to how this exceptional postponement has impacted Tokyo's legacy discourse.

As outlined above, temporality has a range of implications for the concept of legacy. Therefore, this research will take into consideration the temporal aspect of the analysis. The methodological framework will be guided chronologically, in order to provide a comprehensive overview of the concatenation of each of the sub-events that constitute the two analysed Olympics.

2.6 Conclusion: key points from the literature

Based on the literature presented in this chapter, the main theoretical foundations that will underpin the development of the analysis are (1) regionality: events are mediated in different ways depending on the region and nation in which they occur, which deconstructs the 'myth of the mediated centre' (Hepp and Couldry, 2010); (2) discursive events: the concept of 'discourses about an event' is replaced by the concept of 'discursive events' (Hepp and Coldry, 2010); (3) counter-discourses: media events are susceptible to counter-discourses, reinterpretations and reframing of their 'official' narrative; (4) news as cultural products: thinking of ideology as discursive leads to the framing of news as cultural products profoundly affected by authorship (Van Djik, 2008); (5) legacy as narrative: legacy is,

---

[18] Except in 1916 (Berlin), 1940 (Tokyo) and 1944 (London).

therefore, an unstable, momentary set of signs shared, produced and reproduced by individuals (White, 1980); (6) utopian and dystopian legacy: the idea of legacy as a development catalyst promotes 'specific visions of what the desired development should look like' (Girginov, 2018), while on the other hand, the fear of generating white elephants feeds a dystopian narrative of the events (Leopkey and Parent, 2012); and (7) temporality: despite its association with the idea of future projection, the temporality of legacy is expanded to include happenings before, during and after the events (Poynter and MacRury, 2009).

In the next chapter, the methods will be introduced and discussed. As a thesis in Digital Humanities, particular effort was put into the construction of a methodological framework that could be reproducible. Critical thinking on algorithmic black boxes and the limitations of computational methods will also be addressed.

Chapter 3: Methodology

This chapter introduces and gives an overview of the key methods used in this research project.

Like the concept of legacy, Digital Humanities (DH) is a blurred, not-well-defined discipline. Is it the presence of a computer that makes research in the field digital? Is it even a field? A methodology? Or does the word digital perhaps refer to the objects of study, as well as to the methods used?[19] The term 'Digital Humanities' has historically been associated with research applied to digitised material – books, manuscripts, etc. In recent years, however, it is also often associated with analysis conducted on born digital material that emerges in digital forms such as webpages, social media content and so on (Brügger and Laursen, 2019).

Digital here is not understood as the opposite of analogue, as pointed out by Berry and Fagerjord (2017). It is complementary, as it provides researchers with different methodological processes that affect the way the discipline is epistemologically defined. For the authors, 'digital humanities introduce methods that are new to humanities, such as computer statistical analysis, search and retrieval or data visualisation, and apply these techniques to archives and collections that are vastly larger than any researcher or research group can handle comfortably' (Berry and Fagerjord, 2017, p. 9). These 'new' methods may be called digital methods. Rogers (2019) defines digital methods as those dedicated to studying natively digital data, highlighting the difference from culturomics and cultural analytics, whose corpora would be composed of digitised material[20].

---

[19] Those questions and many others were gathered together in *Defining Digital Humanities: A Reader* (Terras, Nyhan & Vanhoutte, 2013). This epistemological exploration assembled a diversity of characteristics that can be used to understand the objectives of a digital humanist.

[20] The Digital Humanities more broadly, however, understands digital methods as methods applied both to the study of born-digital and digitised materials. Alternative nomenclature is 'computational methods'.

Rogers describes the use of digital methods as an 'exploration of how to undertake research with the web, rather than just about it' (2019, p. 17). While these methods have been historically influenced by the work of authors such as Henry Jenkins, focusing on what he called 'online culture' (2006), nowadays they are used to study cultural and social phenomena that occur both online and offline (Rogers, 2019, p 4). Online and offline, however, are not two different social realities but one - online - can be comprehended as a computationally mediated manifestation of the offline. It is this understanding that allows researchers to investigate political phenomena like Brexit by looking at the ways people communicate their thoughts and interact with others in social networks.

Digital Humanities is an interdisciplinary field, as it 'often requires a team of people with diverse backgrounds' (Berry and Fagerjord, 2017, p. 8). This thesis was developed under the CLEOPATRA project – Cross-lingual Event-centric Open Analytics Academy – which brought together an interdisciplinary team of researchers focused on the development of NLP tools to study online content. The design of this thesis's methodology has been highly influenced by the research objectives of this group of scholars.

3.1 Challenging pipelines

Fry (2008, p. 5) conceptualised the process of data exploration into seven steps: 'acquire' is obtaining data; 'parse' is organising the data in a structure; 'filter' is removing what is not relevant; 'mine' is applying statistical methods to the data; 'represent' is showing the data in a graphic form; 'refine' is the improvement of the visualisation to make it clearer; and 'interact' is manipulating the data to allow multiple ways of reading it. In the figure below, Fry (2008, p. 15) represents the interaction among the seven stages of data analysis.

Figure 1 – Pipeline proposed by Fry (2008, p.15) showing the interaction between the seven stages of data analysis
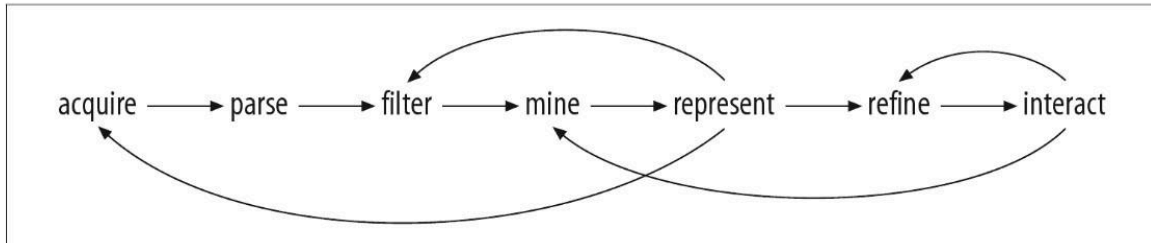
Fry's model is a useful way to structure a data-driven research pipeline. It highlights that the process of exploring data is not as straightforward as his first description of 'stages' would look. This pipeline also places 'representation' not as the end of the process – if it has an end – but as part of the analytical procedures, feeding back into previous stages as a tool to improve the interpretation of data.

What his model does not make clear is that some of these stages might happen simultaneously. The process of gathering information in stage one demands pre-filters and a plan for data organisation. While collecting data, close reading is conducted, which might instigate a rethinking of the data collection itself. Because of this, some of the stages named by Fry will appear in this thesis to describe stages of the process but others will be renamed to improve understanding of the methodology.

I offer below a redesign of Fry's model for data exploration in accordance with this research's pipeline:

Figure 2 – Non-linear pipeline proposed by this project for data exploration
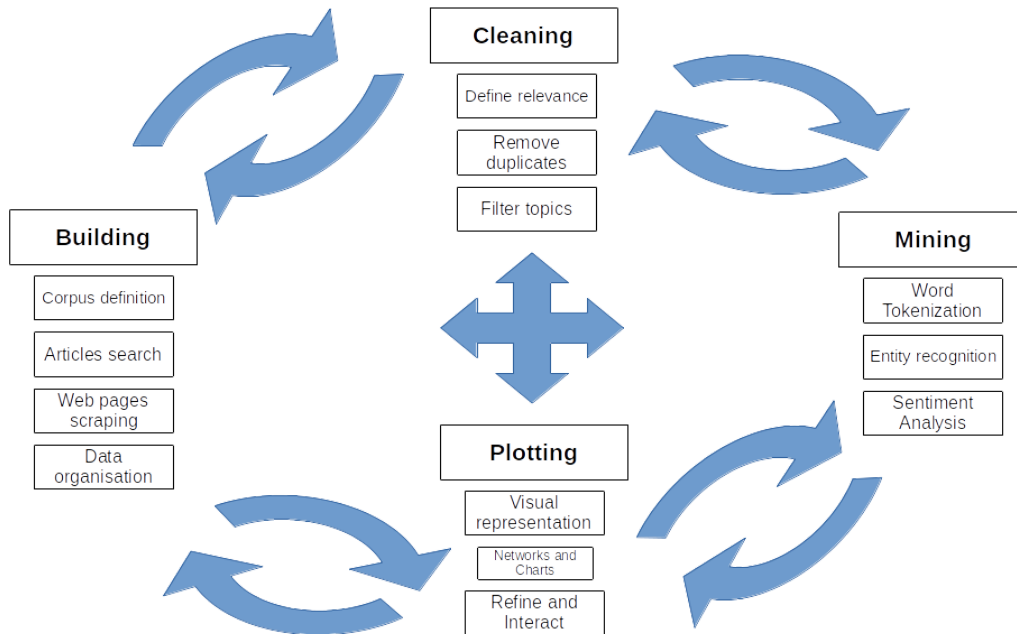
In this diagram, the gathering of data to build the dataset does not occur as the first step of the process necessarily but as part of a constant review of the corpus. By cleaning, mining and plotting, the data had to be altered to answer new questions or simply to become usable when the results were not satisfactory. The following subchapters will describe in detail the methods used in this research to make sense of cross-lingual texts extracted from multiple sources.

3.2 Building the dataset

For the creation of the dataset, I first selected the news outlets that would be relevant for analysis. The key criterion for relevance here is popularity, and I identified the most used websites for news consumption in the United Kingdom and Brazil. I used Google's search engine to search for news articles, opting for advanced search features where I could choose a specific website and delimit important key-words that must appear in the texts. I also collected news articles archived by the UK Web Archive, when available, using the platform

SHINE.[21] The data collection was performed many times during this project, according to the demands that arose throughout the process.

To decide on the outlets whose articles would be integral to this research, I made use of surveys that track the popularity of news outlets in the UK and Brazil. Those surveys were published in 2020, the point of data collection. The popularity of news outlets changes over the years and a decision was made to consider the year of 2020 as the reference point. This means that the articles collected are what users on the internet would find when looking for news articles in this specific period of time.

According to the Ofcom News Consumption Survey 2020 (Ofcom, p. 56) the most accessed websites for news in 2019 in the UK were, in order: BBC, Google, Sky News, The Guardian, The Daily Mail, Google News, Youtube, any local newspaper, Huffington Post, ITV, Buzzfeed, MSN News, The Telegraph, LADbible, Yahoo News, The Independent, The Sun and The Metro. Among these websites, some did not have available data covering the topic 'Olympic legacy', such as MSN News and LADbible. In this case they were removed. Some of them are not media outlets,[22] such as Google, Google News and YouTube. These are news aggregators as they reproduce news produced by news outlets. Although Google was used as the main tool to find news articles in this project, Google platforms, such as Google news and YouTube, were not considered as sources for text analysis.

In some cases, the websites did not allow us to collect their data due to 'page blocks'. This is the case for Huffington Post. When accessing the website, Huffington Post opened a window asking the user to agree with their privacy policies. This window did not allow the scraping

---

[21] It used to be available at https://www.webarchive.org.uk/shine. The interface was taken down by a cyber attack that targeted the British Library in October 2023 and has not been made available anymore until this point in June 2025.

[22] By media outlets we mean media companies specialised in professional news production for newspapers, TV, radio, Internet or any other platform.

algorithm to access the article as it could not click on 'agree' and move on to the next page. The rest of the outlets mentioned in the Ofcom report were included in the research corpus.

Here, I highlight the importance of Google, as it was mentioned three times (Google, Google News and YouTube) by the Ofcom Report. As many people rely on Google to find news, besides its quality outputs for searches, it is a 'place of reference' (Vizeu, 2009) in the sense of being a source of data that frames our access to information. This makes the use of Google's search engine in this research particularly relevant. Rogers (2019) pointed out scholars' criticisms of the use of Google's search engine as its commercial affordances result in biased outputs. However, for the purpose of this research, Google's bias provides insights into the news articles that were most visible and accessible to wider audiences. That is why alternative search engines such as DuckDuckGo.com were not considered.

The report published in 2020 by the Reuters Institute for the Study of Journalism (Digital News Report 2020, p. 90) ranked the most accessed websites for news in Brazil. They are as follows: Globo News (incl. g1.com), UOL online, Record News (incl. R7.com), O Globo, Yahoo News, MSN News, Band News, Folha de S. Paulo, O Estado de S. Paulo, BBC News, regional or local newspaper, Rede TV News, commercial radio news, Buzzfeed News, free city newspaper websites and Jornal Extra. First, it is important to clarify how these media outlets are organised in the country. Globo is a media company with investments in TV, Cable TV, news websites, radio, cinema and streaming on-demand. Because of this, Globo News, O Globo and Jornal Extra share common channels and were clustered together. Band News and Rede TV News were not included because their websites are focused on videos shared by their TV production arm, without texts. The impact of these organisational structures of the Brazilian media will be addressed in more detail in Chapter 5, where the coverage of Rio 2016 is analysed. In both the British and Brazilian corpora, generic references to local newspapers were not considered.

Although this study is focused on news coverage, the method implemented, which combined quantitative exploratory techniques with qualitative analysis, privileged the study of topics

that appeared more frequently in the news articles. Therefore, it was important to bring to the methodology data that allowed for the identification of what was not there: the silences. By bringing in the narratives vocalised by local governments and in activists' blogs,[23] the intention was to identify silences in the media, guiding the analysis not only by what was said but also considering those topics that were ignored or hidden. The choice to include activist blogs is also informed by my interest, as an author, in how civil society responds to mega-events. My understanding as a journalist, which is also confirmed by the analysis in this thesis, is that mainstream media often neglects activists' voices, which are an essential pillar of democratic debates.

Activists' blogs have played an important role in reporting the legacy of the Olympics, particularly as they developed collective observatories to critique the event. For the purpose of this thesis, two official websites of the UK Government (UK Sports and The Queen Elizabeth Olympic Park) and two activist blogs (City Monitor and Games Monitor)[24] were collected. The UK Government official websites were chosen according to their relevance in Google's search ranking, as they appear in the first three pages when searching for 'London Olympic legacy'. City Monitor was found mentioned in the fifth page, when looking specifically for activist blogs in the results. Games Monitor was collected from the UK Web Archive. Games Monitor was the only activist blog covering the Olympics accessible via SHINE, the UK Web Archive platform, and it is not on the live web anymore.

Once these websites were found for the London Olympics, it became important to have similar data sources for Rio, for comparative purposes. For Rio, however, a more active search had to be performed. While for London it was possible to identify some government and activist websites via a general Google search for 'London Olympic legacy', the same did not occur for Rio. For the Brazilian corpus, Arquivo Esporte and Rede do Esporte were

---

[23] Besides their relevance as both appear in the first pages of a Google search for 'London Olympic legacy'.

[24] uksport.gov.uk, queenelizabetholympicpark.co.uk, citymonitor.ai and gamesmonitor.org.uk.

included as official government websites and Rio on Watch as an activist blog.[25] As none of these websites appear in the first pages of Google Search, a more specific search to identify these sources was performed for 'Legado Olimpico ativista blog'[26] and 'Legado Olímpico governo brasileiro'.[27]

3.2.1 Data collection

After selecting the relevant sources, websites from which to extract the articles were identified. The process of collecting texts from websites is commonly referred to as web scraping (Black, 2016). An attempt to scrape news articles directly from the respective outlets' websites using their own search tools was frustrated because most of them relied on Google search engine and when they did not, the search for keywords was not as accurate as Google's. For this reason, a decision was made to use Google as the main tool to find the articles.

To scrape news articles from Google, two python libraries were used: fake-useragent[28] (Kovtun, 2018) and BeautifulSoup[29] (Richardson, 2007). The first requests information from the given website while the second identifies HTML marks[30] and facilitates the identification of specific information such as the pages' title and description. Advanced search was used for

---

[25] arquivo.esporte.gov.br, rededoesporte.gov.br and rioonwatch.org.br.

[26] In English: Olympic legacy activist's blog.

[27] In English: Olympic legacy Brazilian government.

[28] Fake-useragent creates a User Agent to access the websites. Documentation at https://pypi.org/project/user-agents/

[29] Beautiful Soup is a Python library for pulling data out of HTML and XML files. Documentation at https://pypi.org/project/beautifulsoup4/.

[30] An example of HTML mark is '<head>', which indicates the beginning of the metadata information within a html script.

news published by specific outlets by mentioning the domain name before the search for keywords as shown in the example below:

---

site:bbc.co.uk 'olympic legacy' london

'Olympic legacy' was used with quotes to find only results where both words occur together. As Olympic legacy is understood as a concept, using a measure of proximity would not deliver results as accurate as those obtained by looking for the bigram. The output of this process was a list of URLs to be used for text extraction.

The second database used was the UK Web Archive via SHINE, specifically for collecting data for the British corpora. SHINE is a tool to explore data from .uk websites in the Internet Archive between 1996 until April 2013. Although a large part of the content of the UK Web Archive can only be accessed inside the British Library, SHINE is open and its data can be consulted online. While Google proved to be very efficient to find recent news, it was difficult to access old articles published by these outlets via Google search. The UK Web Archive had acquired data from the Internet Archive from thousands of articles about the London Olympics, published by The Guardian, BBC and The Independent between 2004 and 2012. The data is open access and available for download from the SHINE website.

SHINE provides many features for advanced search. For this project, a search for '"Olympic legacy" London' (using double quotes, as required by SHINE for phrase search) was performed. For filtering the outputs, the specific domains this research was interested in extracting were selected: The Guardian, BBC, The Independent and Games Monitor. In SHINE, search results can be downloaded as *.csv* files. This significantly reduced the steps to organise information into a spreadsheet, which was necessary for the Google search. As shown in Figure 3, data provided by SHINE is organised in columns as follows: title of the webpage, date of archiving, archived URL, among other metadata.

Figure 3 – Spreadsheet provided by SHINE for the search '"Olympic Legacy" London'

"Results from the British Library's Shine interface	(webarchive.org.uk/shine)	on 14/11/2020."
"Search Summary: Search Term(s): ""Olympic Legacy"" London	Domain: (Include) ""bbc.co.uk"""	

Title	Date	URL
" ""BBC SPORT | Other Sport | Olympics 2012 | London survives cut"""	05/20/04 12:15 PM	http://web.archive.org/w
BBC SPORT | Other Sport | Olympics 2012 | London survives cut	05/22/04 04:29 AM	http://web.archive.org/w
BBC SPORT | Other Sport | Olympics 2012 | Candidate cities for 2012 0	05/22/04 07:28 AM	http://web.archive.org/w
Layout 1	04/17/05 01:29 PM	http://web.archive.org/w
BBC - Kent - Message board - Kent nightlife	09/01/05 02:04 PM	http://web.archive.org/w
BBC NEWS | Education | Will the Olympics help languages?	09/13/05 02:45 PM	http://web.archive.org/w
BBC SPORT | Other Sport... | Olympics 2012 | London Guide | London's	01/10/06 05:37 AM	http://web.archive.org/w
BBC NEWS | UK | England | London | UK urged to prepare for Olympics	01/16/06 01:32 PM	http://web.archive.org/w
BBC SPORT | Other Sport... | Olympics 2012 | Rivals for 2012: London	03/09/06 03:15 AM	http://web.archive.org/w
BBC NEWS | UK | England | London | Man cleared over Damilola murder	04/05/06 02:08 AM	http://web.archive.org/w
BBC NEWS | UK | England | London | Six nurses fall ill with measles	04/05/06 02:08 AM	http://web.archive.org/w
BBC SPORT | Other Sport... | Olympics 2012 | Coe welcomes Olympic	04/09/06 03:12 AM	http://web.archive.org/w
BBC NEWS | England | London | Blair's plan for Olympic legacy	04/09/06 04:04 AM	http://web.archive.org/w
BBC SPORT | Other Sport... | Olympics 2012 | Coe dismisses £2bn ove	04/10/06 11:59 AM	http://web.archive.org/w

Once all the URLs from Google and UKWA search were gathered, I ran a script to scrape the pages and save their content in a usable file format. I chose to work with *.txt* files as they are easy to manage, store, process and modify. To extract only the relevant text from the web pages - the body text of the articles - I used BoilerPipe[31] (Kohlschütter et al., 2010). This library detects and removes the clutter around the main textual content of a web page. In pages with comments, advertising, links and so on, the use of BoilerPipe prevented this extra information from being mixed with the main article text.[32] By running this script, it was possible to extract and store each article in an individual file.

---

[31] When the methodology for this thesis was designed in 2020, Boilerpipe was a Java library. It was necessary to use a Python wrapper to allow me to use it with Python. The process to adapt the use of this library for Python required several steps: JAVA_HOME environment was required in combination with JPype (a Java-Python bridge). For documentation see https://pypi.org/project/boilerpy3/. However, from 2021 a new version of Boilerpipe was released, updating the codebase to Python, which makes its use much simpler than when the tool was originally used for this study.

[32] In 2022, a comparative evaluation of a diverse range of web scrapers was conducted by the Trafilatura team (https://trafilatura.readthedocs.io/en/latest/evaluation.html). It places Trafilatura as one of the best performing algorithms for scraping data from websites. It seems to be a promising tool to be used in future projects.

3.2.2 Sorry for the inconvenience: the challenge of collecting data from Brazilian websites

The extraction of Brazilian texts in Portuguese was one of the main challenges while scraping news articles. The official page of Rio 2016 has been down since 2018 and the archived version in the Internet Archive does not allow users to access many texts under the section 'News'. This happens due to the way webpages are archived. In many cases, the archive does not collect the entire website. Consequently, it is not possible to navigate through the pages by using the menus or just updating page numbers in the query, as is the case for most live websites. When accessing rio2016.com.br the only information available is the text: 'Sorry for the inconvenience. System is under maintenance and will be back on the 2nd of July, 2018 at 9 a.m.'.[33] The website is one of the many abandoned legacies of Rio.

Collecting data from news sources in Portuguese was also not a simple task. Many websites had a poor structure that was not easily recognizable by the algorithm used to extract texts. Sometimes, non-relevant text (such as menus, or comments) was saved into the document files. Therefore, the process required a one-by-one checking of each of the files produced for each of the specific domains, such as Yahoo News. Non-relevant texts had to be deleted manually from the .*txt* file.[34]

For other websites the situation was even worse. Websites like R7.com could not be read by the algorithm at all, returning a text file with only code rather than the body text of the article, as is shown in the image below.

---

[33] This was true in May 2020. Now, in March 2023, the website was completely taken down and no content is shown when accessed.

[34] Non-relevant text here means any form of text that is not part of the news article, such as comments produced by the audience or advertisement, for example.

Figure 4 – Text file produced with BoilerPipe from R7.com

```
 1 .logo-r7-svg{
 2      fill: #fff;
 3      }
 4      .show .logo-r7-svg{
 5      fill: #4565af !important;
 6      }
 7 r7adv.incrementUnit({
 8      code: 'vertical-header-ads',
 9      sizes: [[728, 90], [970, 90], [970, 250], [970, 150], [920, 100]],
10      });
11
12      r7adv.defineAd({
13      sizes: [[728, 90], [970, 90], [970, 250], [970, 150], [920, 100]],
14      id: 'vertical-header-ads',
15      pos: 'header'
16      });
17 window.googletag = window.googletag || {cmd: []};
18      googletag.cmd.push(function() {
19      googletag
20        .defineSlot(window.r7ad.hash, [1, 1], 'div-gpt-ad-1642428980255-0')
21        .addService(googletag.pubads())
22         .setTargeting('pos', 'in-image');
23
24      googletag.pubads().setTargeting('pageId', window.location.href);
25      googletag.enableServices();
26      });
27 googletag.cmd.push(function() {
28      googletag.display('div-gpt-ad-1642428980255-0');
29      });
30 .logo-r7-svg{
31      fill: #fff;
32      }
33      .show .logo-r7-svg{
34      fill: #4565af !important;
35      }
36 Para ver este vídeo, ative o JavaScript ou considere uma atualização para um navegador da Web que suporte vídeo HTML5
37 window._taboola = window._taboola || [];
38
39      _taboola.push({
40      mode: 'thumbs-feed-02-c',
41      container: 'taboola-below-article-thumbnails',
42      placement: 'Below Article Thumbnails',
43      target_type: 'mix'
44      });
45 var box_key = "publicidade-52e2c2a06c4db2bfa3003fe6";
46
47  (function(){
48
49    var config = {
50      code: 'box-publicidade-island-ad-52e2c2a06c4db2bfa3003fe6',
51      sizes: [300, 250],
52    }
53
```

Collecting data from Brazilian newspapers for the purpose of this research was also challenging because of other factors such as the concentration of media ownership in Brazil, which strongly affects the variety of content produced in the country. This specific challenge as well as those related to the lack of tools available for multilingual analysis are further described and discussed in more detail in Chapter 5. It is worth mentioning here also the lack of a Brazilian web archive or a tool like SHINE for Brazilian websites.

3.2.3 Cleaning and filtering the data

The cleaning process consisted of ensuring that all the scraped articles were both unique and relevant for analysis, i.e. they covered the London and/or Rio Olympics. The first step was removing duplicates. As many articles, especially those collected from the UKWA, appeared more than once in the dataset, it was necessary to deduplicate using a combination of different techniques. I used the deduplication tool available in LibreOffice Calc[35] and also compared text snippets to detect repetition using Python.

Although I specified the topic when searching on Google and SHINE, both tools provided inaccurate outputs - of course in different proportions. In many cases, the searched word appeared in peripheral locations on the web page – such as menus, comments, adverts – and not in the body text. As a result, once BoilerPipe had removed the peripheral information, there was no longer any mention in the file of the word 'legacy'. To identify these issues, I used a Jupyter notebook[36] to filter the text files, requesting only those mentioning the word 'legacy'[37] (Appendix C). They were then saved in a new folder as 'Olympic content'. The following figure[38] shows the percentage of articles remaining in the dataset after applying the filter.

---

[35] In the case of spreadsheets produced with SHINE, it was possible to simply identify repetition in the news titles and/or their original URLs. The archived URLs, those generated by the UK Web Archive, cannot be used for deduplication as they are always unique, even though the same page was collected many times. For Google search, the comparison of text snippets was the best approach. However, duplicate results were much rarer.

[36] Jupyter Notebooks is an open-source web application to create and share codes.

[37] In English: Legacy, legacy, Legacies, legacies. In Portuguese: legado, Legado, legados, Legados.

[38] This chart was produced with RAW Graphs (Mauri et al. 2017), an open-source data visualisation framework. The Sankey diagram was chosen to illustrate the flow of resources, gathering news outlets according to their source and number of articles. RAW Graphs is maintained by the DensityDesign Research Lab, Politecnico di Milano University.

Figure 5 – Percentage of remaining articles after filtering. In blue, those scraped from the UKWA and in orange, from Google
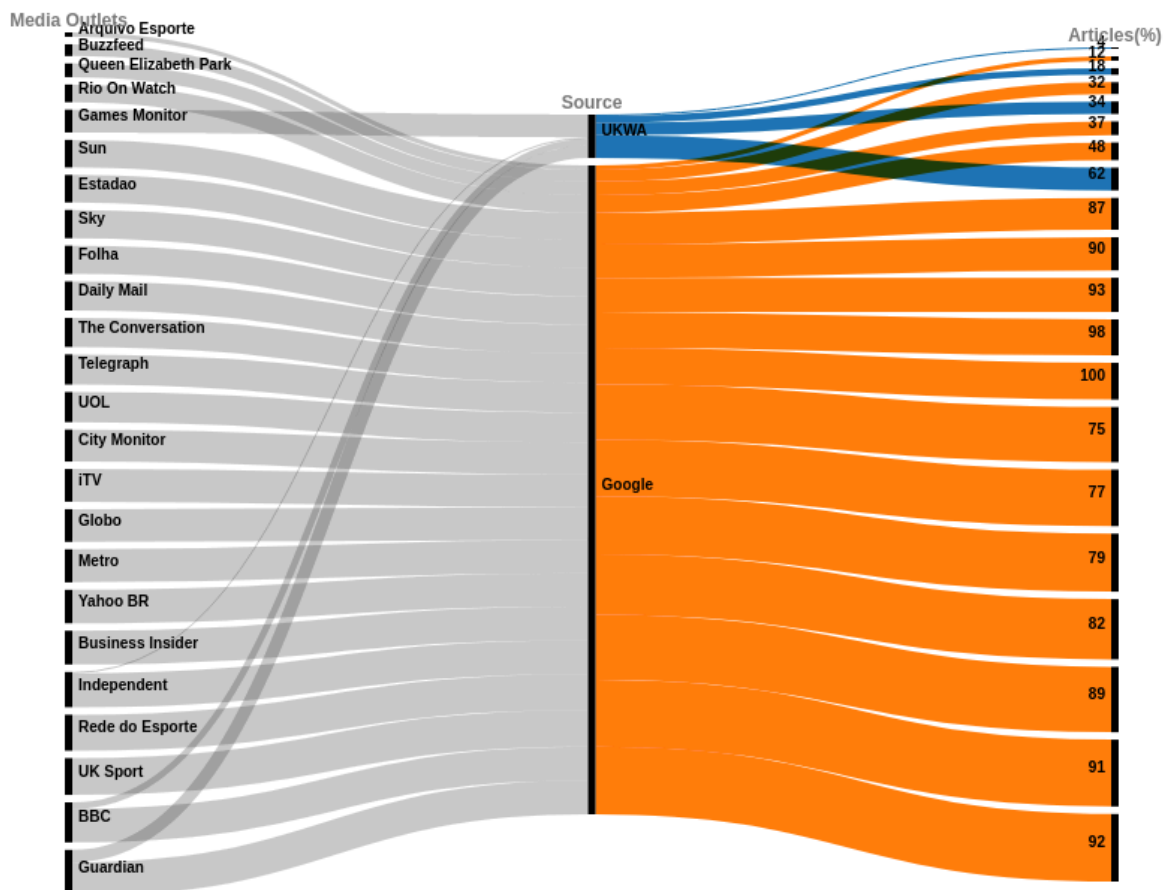
The chart shows in blue the articles extracted from the UKWA using SHINE; in orange, those from Google. The blue lines shed light on the small amount of relevant articles scraped from the UKWA. This was due to the fact that Google provided, in most cases, results where the word searched had relevance on the page – it usually appeared in the title or main body text. This was not the case for the data acquired using SHINE. The tool looked for the keywords in the entire archived document, returning irrelevant outputs. For example, in the spreadsheet containing outputs for the search '"olympic legacy" London Domain: "uksport.gov.uk"', 19,899 duplicates were identified and removed, leaving only 103 unique values in the folder.

3.2.4 Web archives as sources for quantitative research

The use of the UK Web Archive as a source raises important questions regarding the nature of its data and exploratory potential. As pointed out by Brugger (2018, p.74) 'efficient scholarly use of the archived web presupposes a detailed understanding of the digitality of the web and its changes when collected, preserved and made available for research purposes'. Winters (2019, p. 83) highlighted aspects of web archives that differ from other sources researchers are more familiar with: archived sources vary from other archived sources but also from the live web; although born digital, access to them can be physically restricted; the particularities of their duplicates and overlaps are difficult to identify; and they cross territorial borders although national and institutional limits apply.

When working in partnership with the British Library during a secondment, I worked on the UK Web Archive *in loco* for two months. Although I was able to read several news articles archived in special collections within the UK Web Archive, for the purpose of this study, this access was inefficient. Besides limits imposed in terms of access, copies and reproduction of archived data, one of the biggest barriers was the limitation of tools to search through the archive. As it was not possible to download the data and build my own search engine, as a researcher I was restricted by the UKWA websites' affordances. Overall, transforming archived news articles into manageable data for quantitative analysis tends to be arduous, as this process challenges the organisational logic of web archives in which users are looking for specific pages for close reading, like visitors to traditional archives.

The solution found to tackle these limitations was to rely on SHINE and its publicly available data. Although providing users with free access, SHINE's data also imposed constraints such as those previously mentioned: excess of duplicates and lack of accuracy in the search for topics. For example, SHINE provides in the metadata the date of archiving but not the date the original page – or news article – was published. Therefore, I had to extract such information from the HTML of each page.

3.3 Overview of the data, collection methods and storage

In this section, I provide some visualisations to facilitate an understanding of the processes and the data discussed above. In Figure 6, there is a scheme describing the processes undertaken to identify the data sources, and collect, clean and filter the data. In sum, first the sources for data collection were identified based on their popularity. Data retrieved from Google was complemented with data collected from the UK Web Archive. On Google, searches for '"Olympic legacy" London' (in English) and '"Legado Olímpico" Rio' (in Portuguese) were performed. To collect the data, a python script using Fake-useragent and Beautiful soup was first used to get and create a list of URLs for the news articles presented through Google's outputs. For SHINE, a search was performed and the platform provided the URLs organised in a spreadsheet for download.

With the list of URLs, BoilerPipe was used to retrieve only the body text of the news articles, ignoring any peripheral information that the pages might contain. For cleaning and filtering, a duplication removal was performed using LibreOffice Calc as well as text snippets comparison. At the end, files were reviewed to check whether the word 'legacy' was still present after the removal of clutter.

Figure 6 – Processes for gathering and preparing data

Processes for gathering and preparing the data
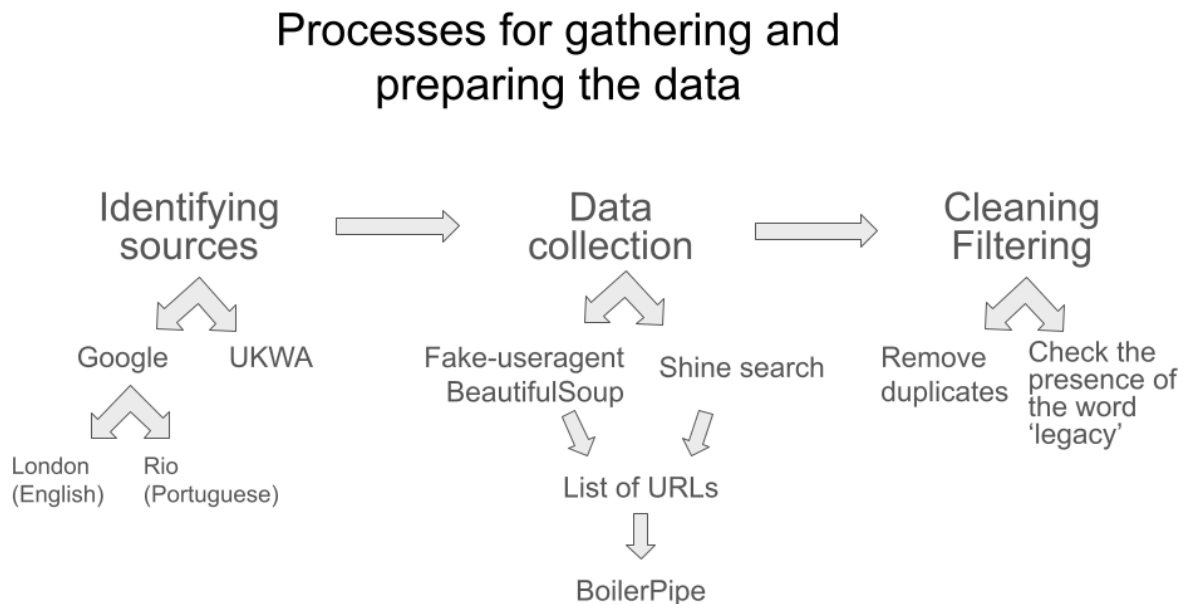
The data storage scheme with its multiple directories is represented below in Figure 7. Inside the folder Data there are two other folders named London and Rio, which contain documents related to each of these events. For London, there were two sources of information, Google and UKWA; for Rio, just Google. Each source has three categories of articles: news, government and activist. Inside each of these categories are the folders named according to the media outlets, for example BBC. Under the media outlet, there is one folder containing the raw article, named full-articles, and one containing the filtered and cleaned documents, named olympic-articles.

Figure 7 – News articles storage scheme

```
NewsArticlesStorage
├─ Data
└─ Event (Eg. London)
   ├─ Google
   │    ├─ News
   │    │    └─ News outlet (Eg. BBC)
   │    │         ├─ full-articles
   │    │         └─ olympic-articles
   │    ├─ Government
   │    │    └─ Goverment webiste (Eg. UKSport)
   │    │         ├─ full-articles
   │    │         └─ olympic-articles
   │    └─ Activist
   │         └─ Activist blog (Eg. Games Monitor)
   │              ├─ full-articles
   │              └─ olympic-articles
   └─ UKWA
        └─ Domain (Eg. Guardian)
             ├─ full-articles
             └─ olympic-articles
```

In Figure 8 below, there is a scheme to illustrate the final dataset with more detailed information. The text files are stored according to category, event, language, source and outlet. The outlet with the highest number of articles in this dataset is The Guardian, followed by Globo, BBC, The Telegraph and Daily Mail. The final number of articles in the dataset is 2382.

The figure highlights the scarce amount of data in Portuguese (orange), in comparison to English (blue). There is a discrepancy in the availability of news articles from the above mentioned outlets.

This dataset was used to answer most of the research questions proposed in this thesis. However, some of the questions required a more precise and curated dataset for best results. For this purpose, a sub-dataset was created and is described in the next section. The main application of this sub-dataset was in the Sentiment Analysis, which is further explained in section 3.7 of this chapter.

Figure 8 – Dataset overview

Dataset
overview
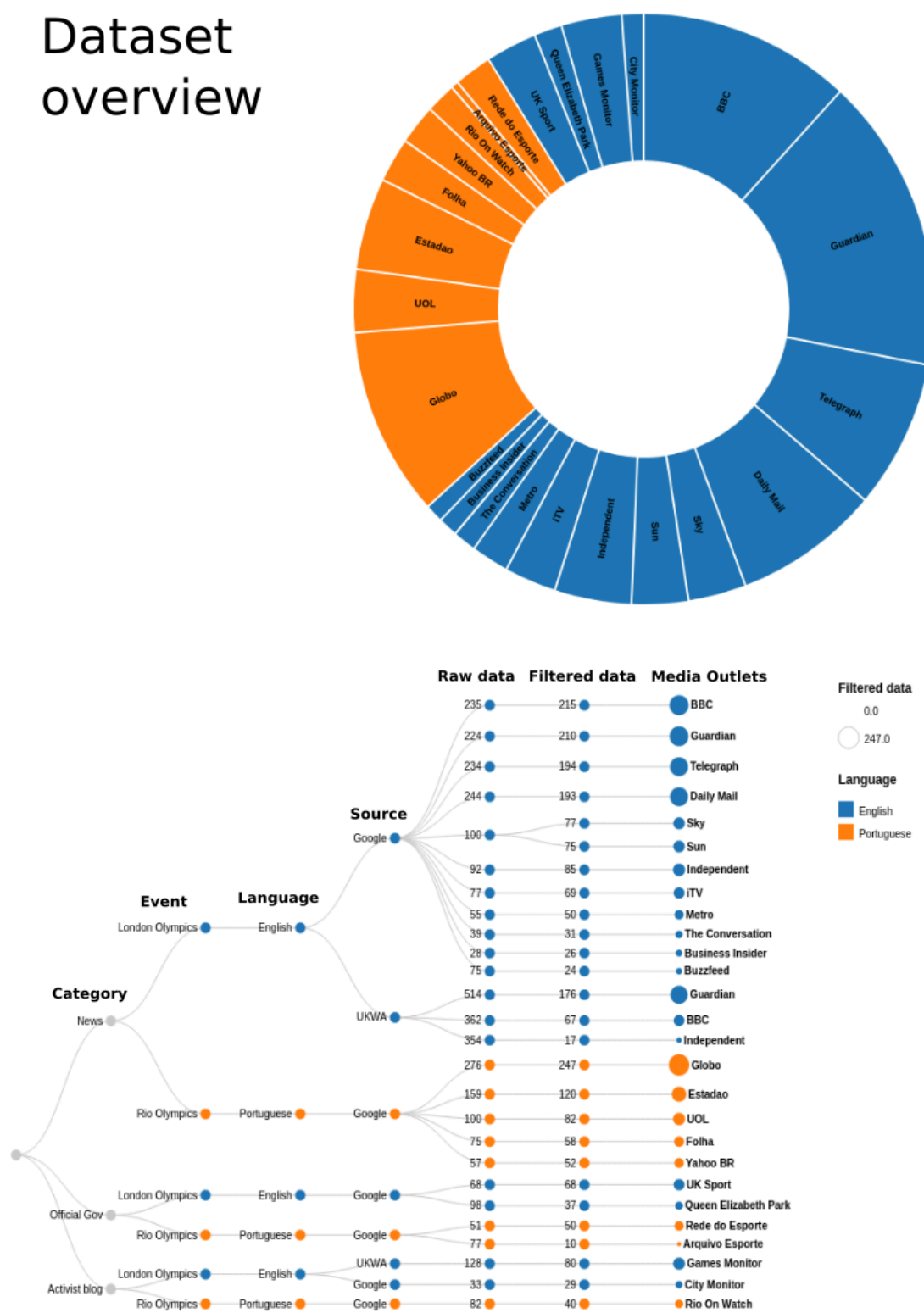
3.4 Elaborating a curated sub-dataset for deep analysis

One of the objectives of this thesis is to conduct a comparative analysis of the news media coverage of both Olympic legacies by British and Brazilian outlets. Initially, there was a plan to scrape news articles published by the British press covering London and Rio and the same for the Brazilian press. However, there was not a significant amount of articles about Rio's legacy in some of the British websites and about London's in some of the Brazilian websites. Moreover, there was a lot of uncertainty with regards to whether articles were specifically covering the event of interest as the main topic. Where an article mentions in the body text both 'London Olympic legacy' and 'Rio Olympic legacy', it is difficult to affirm which or if one of the events is the main theme of the text.

To facilitate computational analysis, where precision on the main topic discussed in the news article is necessary, a deep curation of a sub-dataset was performed. For the United Kingdom, three newspaper websites (guardian.co.uk, telegraph.co.uk and dailymail.co.uk) and one broadcast media website (bbc.co.uk) were selected. For Brazil, three newspapers websites (oglobo.globo.com.br, Estadao.com.br and Folha.uol.com.br) and one broadcast media website (g1.globo.com) were also chosen.

In a spreadsheet (see Figure 9), each article was labelled with the city it reported as the main topic. Besides London and Rio, articles covering other cities like Barcelona and Athens were found and removed from this sub-dataset. At the end, the text files were reorganised according to which city they covered and which country the publisher is originally from. This way, it was possible to perform an analysis to explore in more depth how the British media reported on Rio's legacy and how Brazilian media reported on London's legacy. This also allowed for the comparison of internal and external media coverage of these major events.

Figure 9 – Spreadsheet for data filtering according to narrative relevance

	A	B	C	D	E	F	G	H	I	J	K
	rank	date	title	link	rio	london	olympic	legacy	olympic legacy	verify	other city
	119	03/08/2015	Olympic chiefs to order testing	https://www.theguardian.com/sport/20	10	0	14	1	1	Rio	
	7	04/08/2015	The Rio property developer ho	https://www.theguardian.com/sport/20	26	2	36	3	1	Rio	
	43	04/08/2015	Rio Olympics: public scepticisr	https://www.theguardian.com/sport/20	19	1	20	3	1	Rio	
	40	04/08/2015	Rio Olympics: public scepticisr	https://www.theguardian.com/sport/20	19	1	20	3	1	Rio	
	115	04/08/2015	Government launches consulta	https://www.theguardian.com/uk-news/	1	6	6	1	1	London	
	157	05/08/2015	Green favelas? The promise of	https://www.theguardian.com/world/vie	2	0	2	1	1	Rio	
	8	05/08/2015	Sustainable slums? The promis	https://www.theguardian.com/environn	5	0	6	3	2	Rio	
	206	22/08/2015	Bird's Nest stands as an empty	https://www.theguardian.com/sport/blc	0	1	9	0	0	DELETE	
	16	28/10/2015	Brazil officials evict families fr	https://www.theguardian.com/world/20	10	0	12	1	1	Rio	
	36	08/12/2015	Rio Olympics linked to widesp	https://www.theguardian.com/world/20	6	0	13	1	1	Rio	
	235	09/12/2015	Hamburg - The Guardian	https://www.theguardian.com/cities/20	0	2	14	1	1	other	Hamburg
	77	05/02/2016	Zika crisis and economic woes	https://www.theguardian.com/world/20	15	0	22	2	1	Rio	
	56	21/04/2016	Deaths on collapsed Rio de Jai	https://www.theguardian.com/world/20	7	0	9	2	1	Rio	
	211	18/05/2016	No one wants Manchester Un	https://www.theguardian.com/football/	3	0	8	2	1	Rio	
	48	17/06/2016	Rio de Janeiro governor declar	https://www.theguardian.com/world/20	10	0	8	0	0	Rio	
	13	11/07/2016	Rio mayor Eduardo Paes: 'The	https://www.theguardian.com/cities/20	17	1	22	3	1	Rio	
	70	15/07/2016	The Games: A Global History c	https://www.theguardian.com/books/20	2	3	22	0	0	DELETE	
	52	16/07/2016	Will the Olympics offer Brazil a	https://www.theguardian.com/world/20	19	2	27	1	1	Rio	
	34	27/07/2016	Revealed: the biggest threat t	https://www.theguardian.com/sport/20	3	1	23	0	0	Rio	
	29	27/07/2016	Revealed: the biggest threat t	https://www.theguardian.com/sport/20	3	1	23	0	0	Rio	
	87	27/07/2016	The mixed fortunes of Melbou	https://www.theguardian.com/sport/20	3	2	47	1	1	Rio	
	24	29/07/2016	The trouble with the Olympic	https://www.theguardian.com/sport/20	7	16	35	2	0	Rio	
	17	29/07/2016	Students Speak: the Rio Olymp	https://www.theguardian.com/global-de	24	1	21	2	1	Rio	
	67	29/07/2016	Olympic legacy alive and well	https://www.theguardian.com/society/2	1	2	4	3	3	London	
	55	30/07/2016	Sport's tarnished image create	https://www.theguardian.com/sport/blc	2	2	8	2	2	Rio	
	3	02/08/2016	'The only Olympic legacy I see	https://www.theguardian.com/global-de	11	0	22	3	1	Rio	
	15	02/08/2016	Rio Olympics: view from the fa	https://www.theguardian.com/global-de	4	0	9	2	0	Rio	
	10	02/08/2016	Olympic Games 2016: how Ric	https://www.theguardian.com/sport/20	19	0	18	8	2	Rio	
	19	03/08/2016	Why is Rio de Janeiro finding i	https://www.theguardian.com/global-de	13	0	8	2	1	Rio	
	99	03/08/2016	Brazil deploys over 1,000 troo	https://www.theguardian.com/world/20	7	0	6	1	1	Rio	

rio_theguardian_verify

The spreadsheet is divided into 11 columns showing: (A) position of the news article in google ranking; (B) date of publication; (C) title of the news article; (D) URL for the news article; (E) number of mentions of Rio; (F) number of mentions of London; (G) number of mentions of the word 'Olympic'; (H) number of mentions of the word 'legacy'; (I) number of mentions of the bigram 'olympic legacy'; (J) manual classification made by the author, assigning labels according to the main event covered by the article; (K) manual classification made by the author for events that are not the London and Rio Olympics.

As the storage and composition of this sub-dataset differs from the main dataset presented in the section above, a new scheme was designed to present its directories (see Figure 10).

Figure 10 – Sub-dataset storage scheme

```
Sub-DatasetStorage
├─ Data
└─ Event (Eg. London)
    ├─ English
    │    └─ Brisith News outlet (Eg. BBC)
    │         ├─ full-articles
    │         ├─ olympic-articles
    │         └─ curated-articles
    └─ Portuguese
         └─ Brazilian News outlet (Eg. Globo)
              ├─ full-articles
              ├─ olympic-articles
              └─ curated-articles
```

Inside the main folder 'Data' is the folder 'Event', composed of two folders 'English', for articles published in English language, and 'Portuguese', for articles published in 'Portuguese language'. Inside each of them, there is a folder for each news outlet, which comprises three other folders: 'full-articles', for the raw data; 'olympic-articles', for filtered articles mentioning the word 'olympics'; and 'curated-articles', for those confirmed by the author as belonging to either the London or Rio clusters.

In Figure 11, it is possible to visualise the distribution of articles according to each of the different outlets. It shows the event covered by the articles (London 2012 or Rio 2016); the media outlet; and the number of articles published by each outlet.

Figure 11 – Distribution of news articles in the sub-dataset
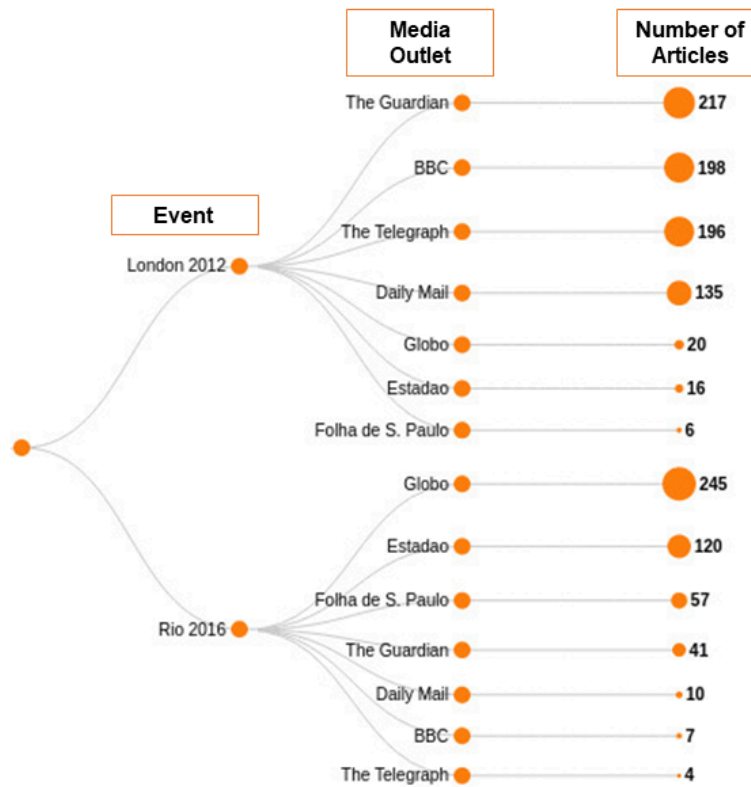
The total number of articles in the sub-dataset is 1271, divided into 464 articles in Portuguese and 807 articles in English. The outlets with the most articles are The Guardian, covering London 2012 (217), and Globo, covering Rio 2016 (245). The outlets with the fewest articles are Folha de São Paulo, covering London 2012 (6) and The Telegraph, covering Rio 2016 (4).

The number of articles in the Brazilian Media Coverage of London Corpus and British Media Coverage of Rio Corpus is smaller than media outlets covering events hosted by their own nation. These discrepancies and the impact they have on the analysis are further investigated and discussed in Chapter 6.

The different corpora produced from this data collection were named as follows: British News Media Corpus, British Government Corpus, British Activist Corpus, Brazilian News Media Corpus, Brazilian Government Corpus, Brazilian Activist Corpus, Brazilian Media

Coverage of London Corpus, and British Media Coverage of Rio Corpus. A brief description is provided below:

- British News Media Corpus: composed of articles curated for the sub-dataset published by British Media covering the London 2012 Olympics.

- British Government Corpus: composed of articles published by the British government bodies websites.

- British Activist Corpus: composed of articles published by activist websites on the London 2012 Olympics.

- Brazilian News Media Corpus: composed of articles curated for the sub-dataset published by the Brazilian Media covering the Rio 2016 Olympics.

- Brazilian Government Corpus: composed of articles published by the Brazilian government bodies websites.

- Brazilian Activist Corpus: composed of articles published by activist websites on the Rio 2016 Olympics.

- British Media Coverage of Rio Corpus: composed of articles curated for the sub-dataset published by the British Media covering the Rio 2016 Olympics.

- Brazilian Media Coverage of London Corpus: composed of articles curated for the sub-dataset published by the Brazilian Media covering the London 2012 Olympics.

Throughout the thesis, I refer to those corpora to facilitate precision on which data was taken into account in the analysis.

3.5 Mapping complex territories

Early research on digital material, and particularly on the Internet, made use of traditional disciplines and methods, such as ethnography or focus-group, to investigate social phenomena occurring in or mediated by electronic media (Venturini, 2012). Although those methods remain relevant for the development of digital research, there are some particularities that are intrinsic to digital mediation. If, according to Latour (2007, p.71) 'hitting a nail with and without a hammer, boiling water with and without a kettle' are not the same thing, the presence of tools such as computers has an impact on the activities performed by humans. Based on this understanding, Latour (2007) developed the Actor-Network Theory, in which everything that makes a difference to a phenomenon must be recognised as an actor. An Actor, then, is not just a human attribute; it can characterise animals, tools or anything else whose presence or absence is responsible for provoking changes. One of the developments of Actor-Network Theory, or simply ANT, is the Cartography of Controversies (Latour, 2007).

Latour's Cartography of Controversies arose as an attempt to observe and represent complex computer-mediated debates such as those related to climate change, where it is possible to identify dissonant ideas being perpetuated by a diversity of actors. Controversies are, then, especially characterised by the lack of concordance or definition. This is also what makes it challenging to explore the concept of legacy, and as happens in debates about global warming, the dissonance is not always explicit. Disputes about the meaning of Olympic legacy emerge as a consequence of multiple interests in qualifying or disqualifying the benefits or harms that result from hosting the event. As in Latour's controversies, Olympic legacy is at the centre of a negotiation between actors interested in making sense of the term. It is also dynamic as it is open to gathering new meanings through the different uses that actors make of it.

This non-static property is one of the main reasons why the fundamentals of cartography were identified as an effective approach for this study. According to Venturini (2010),

mapping cannot tempt us to simplify the complexity of social phenomena. The author used the metaphor of a magmatic flow to refer to the instability of this chaotic scenario (2010, p. 263). Venturini offers a step-by-step method for the exploration of controversies, which he called the five observation lenses. The first one highlights the importance of mapping references in order to identify how the literature articulates the different discourses. The second invites the researcher to question who the actors are. The third emphasises that more than looking at the objects, the researcher must look at the movement, the flow and the changes. In the fourth lens, Venturini introduced the concept of cosmos, which focuses on the actors' interests and therefore relates to how ideologies are organised on the web. The last layer refers to what the author called cosmopolitics: the belief that there is no objective reality to be proven. It means that by observing the different arguments, the objective of the research is not finding a winner but describing all the multiple aspects. This idea, however, does not imply a positivist quest for neutrality. It actually helps us to understand the process not as a judgement per se but as an effort to make the complexity of reality a bit more legible.

Throughout this thesis, Venturini's five observation lenses were used as a guide for a critical construction of the methodology. In this way, the study started by looking at how the concept of Olympic legacy has been described and analysed by authors from different disciplines, such as history, urban development and economics. Second, in the analysis in Chapters 4 and 5, there was an effort to identify the actors – or entities – that are central to the narrative. They can be decision-makers in politics, international organisations and companies; activists and local citizens; or even things such as the Olympic stadium in London or the Olympic park in Rio.

The third point is central for analysis: as this is a study of narrative, emphasis was given to the actors-objects movement. Questions were developed to guide the analysis in the direction of identifying how actors emerged in the story, what their influences in the event were and, sometimes, how they left and returned to the discussion over time. This careful look at movements inevitably directed the analysis towards critical thinking in relation to actors' interests. At this point, besides the information provided by the news articles analysed and the

secondary sources, it was important to look at the extra-textual elements – also referred to as inter-discourses – such as the authors of the articles and the newspapers' economic interests. For example, I identified (see Chapter 5 for more details) that an article published by The Guardian in 2012 under the section 'News', entitled 'The Olympic legacy: creating a new community for London in Stratford' (Meredith, 2012), was written by Dale Meredith, executive director of Triathlon Homes, a company with investments in buildings constructed for the Olympic village in London. The buildings are often accused by the local community of being unaffordable and of having impacted housing prices in Stratford.

The last point made by Venturini refers to 'cosmopolitics'. For the author, this demands abandoning 'the belief that behind all ideologies and controversies, some objective reality must exist independently from what actors think or say' (2010, p. 267). This understanding guided the analysis towards the consideration of a multiplicity of meanings for the concept of legacy, avoiding the attempt to simplify a complex dispute involving different actors. On the dispute aspect, the author also highlighted that:

> Sometimes disputes are temporarily silenced by the fact that some cosmos has prevailed over the others or by the fact that actors have found a resisting compromise, but no agreement, no convention, no collective reality has ever come without discussion. This does not mean that we could never inhabit a peaceful world, that we could never align our visions, that we could never agree on truth. (Venturini, 2010, p.267).

The objective is then to look at the tension – at the dispute – in order to produce an interpretation of the discourses. Tension and dispute, as discussed above, are not merely textually expressed but also constituted by silences.

3.5.1 *Ceci n'est pas une pipe*: drawing representations

As pointed out by Venturini (2012), exploration and representation cannot be dissociated from each other in cartography. Mapping involves drawing a representation. The map is not

therefore the observed territory but a strategy to make the territory and its nuances and complexities legible. As in Magritte's famous painting, The Treachery of Images (1929), what can be seen is not a pipe but rather an image of a pipe. This concept has been key to guiding the methodological construction of this study.

In order to produce these representations, it is important to observe Venturini's (2012) precautionary warnings, divided into three points: adaptation, redundancy and flexibility. First, the author alluded to the image of constructing a building to state that the construction must be as adapted to the irregularities of the terrain as the map is to what it intends to represent. In describing redundancy, Venturini highlighted the importance of producing many maps representing different perspectives of the phenomenon, as a single map could not be exhaustive and legible at the same time. The third point is flexibility, which means that the exercise of reducing complexity in order to make the map readable should not limit the possibilities of its use. This means that the researcher should provide readers with the possibility of making new connections and explorations through the map: 'What is true for buildings is true for representations as well: the better they are built (the more they adapt to their territory, the more they are redundant and flexible), the more solid they will be' (Venturini, 2012, p.8).

Having described some of the assumptions associated with the cartographic method, this study will make use of it by aiming to avoid reducing and simplifying the multiple dimensions and aspects of Olympic legacy. The purpose is to produce observations and representations that draw out the nuances involved, understanding not only the objects but the methodological construction itself as a subjective process. Working with such fluid objects requires the study to be flexible and open to changes and adaptations over time. Strategies change according to the results we get applying one or another method, and these processes will be described in detail throughout this thesis.

One of the methods that have been used in this thesis is network graphs. While these have been widely studied, it is only recently that researchers have begun to pay attention to their

narrative and storytelling potential (Bounegru et al., 2017, p. 700). Venturini et. al. (2017, p.11) described how visual properties – such as 'topology, density of connections, absence of connections, size, position and colour of nodes' – can be used to construct narrative meaning through network graphs. The visual representation of narratives using networks is an effective way to explore data in this corpus, as it allows researchers to look at the objects of study through different frameworks. Venturini et. al. (2017), analysing the narrative of Iliad, divided these frameworks into three categories: panorama, vantage and journey. 'Panorama' refers to the overview of the narrative, where the reader can identify clusters, groups and their territorial distribution. 'Vantage' refers to the importance of the position that each node - actor - occupies in the network. 'Journey', in turn, refers to the path connecting different nodes. These variants of possible ways to read a network will be used to design an analysis that identifies these characteristics in the corpora, in order to represent the media narrative of the Olympic legacy.

3.5.2 Identifying entities through networks

To produce a network representation of entities mentioned in the news articles, different techniques of named entity recognition (NER) were combined. First, Spacy was used to recognise the entities and label them according to their type (for example: person, organisation, date). However, Spacy cannot differentiate between two people with the same surname based on context. By combining Spacy with Wikifier (Brank, Leban and Grobelnik, 2017), I aimed to disambiguate the entities. Wikifier[39] is a semantic annotation service that links concepts to Wikipedia pages. This way, it provides each entity with a unique identifier that can be used to relate the words to their correct entity. It worked in most of the cases in this thesis, with a few easily identifiable mistakes for English and significantly more issues for Portuguese – these issues will be addressed in the analysis. The nodes spreadsheet was

---

[39] The code used to link the entities provided by Spacy with Wikifier can be found here: https://github.com/cleopatra-itn/news-cartography-analysis

created with the Wikipedia IDs and their labels – the entities. In spreadsheet A, there are two columns: (A) composed of Wikipedia IDs, and (B) stating the name of the entity.

Figure 12 – Sample of nodes and edges spreadsheets. Spreadsheet A (nodes) connects IDs to their labels. Spreadsheet B (edges) connects source and targets, attributing a weight to the edges

	A	B	C	D
1	ID	Label		
2	Q712089	Luís Alves de Lima e Silva, Duke of Caxias		
3	Q23437074	Ministerio		
4	Q24255536	Future Arena		
5	Q198032	Esporte Clube Bahia		
6	Q10317356	Leonardo Picciani		
7	Q41428	Rio de Janeiro (state)		
8	Q6756538	Marcelo Crivella		
9	Q61	Washington, D.C.		
10	Q804016	Baixada Fluminense		
11	Q8613	2016 Summer Olympics		
12	Q5013660	CPLP Games		
13	Q1439498	Fot		
14	Q59099543	Técnicos em Acampamento - 898		
15	Q10330418	Ministry of Sports (Brazil)		
16	Q12705	Renewable energy		
17	Q8678	Rio de Janeiro		
18	Q98792896	Picciani		
19	Q37070629	Crivella		
20	Q513114	Patrícia Amorim		
21	Q5512664	G1 (website)		

	A	B	C	D
1	Source	Target	Type	Weight
2	Q804016	Q6756538	Directed	4
3	Q804016	Q10330418	Directed	3
4	Q804016	Q513114	Directed	3
5	Q804016	Q23437074	Directed	1
6	Q804016	Q5013660	Directed	5
7	Q804016	Q12705	Directed	1
8	Q804016	Q61	Directed	1
9	Q804016	Q41428	Directed	5
10	Q804016	Q1439498	Directed	4
11	Q804016	Q712089	Directed	3
12	Q804016	Q24255536	Directed	5
13	Q804016	Q59099543	Directed	1
14	Q804016	Q37070629	Directed	4
15	Q804016	Q8613	Directed	6
16	Q804016	Q10317356	Directed	1
17	Q804016	Q8678	Directed	10
18	Q804016	Q98792896	Directed	1
19	Q804016	Q198032	Directed	4
20	Q6756538	Q10330418	Directed	11
21	Q6756538	Q513114	Directed	8

Spreadsheet A   Spreadsheet B

In Spreadsheet B (above), the extracted entities were then connected in a relation 'source-target' as they appeared in the same document, to form an edge between entities. A weight was attributed based on the frequency[40] with which two entities are mentioned together in one article. Spreadsheet B, therefore, is composed of four columns: (A) Source; (B) Target; (C) Type; and (D) Weight. Both spreadsheets were loaded into Gephi (Bastian, Heymann and Jacomy, 2009) and a network was produced. The layout of the nodes of the network was produced using the ForceAtlas2 algorithm (Jacomy et al., 2014). This

---

[40] The number used as 'weight' is the number of times both entities were mentioned in the same document.

force-directed algorithm spatialises the network, making nodes repel from each other, while the weighted edges attract them back (Jacomy et al., 2014). To identify modules – strong divisions in the network – I applied Gephi's modularity algorithm for clustering layout. This can be seen through the use of different colours in the example below:

Figure 13 – Network visualisation of entities mentioned in news articles about the Olympic legacy of London by the British press
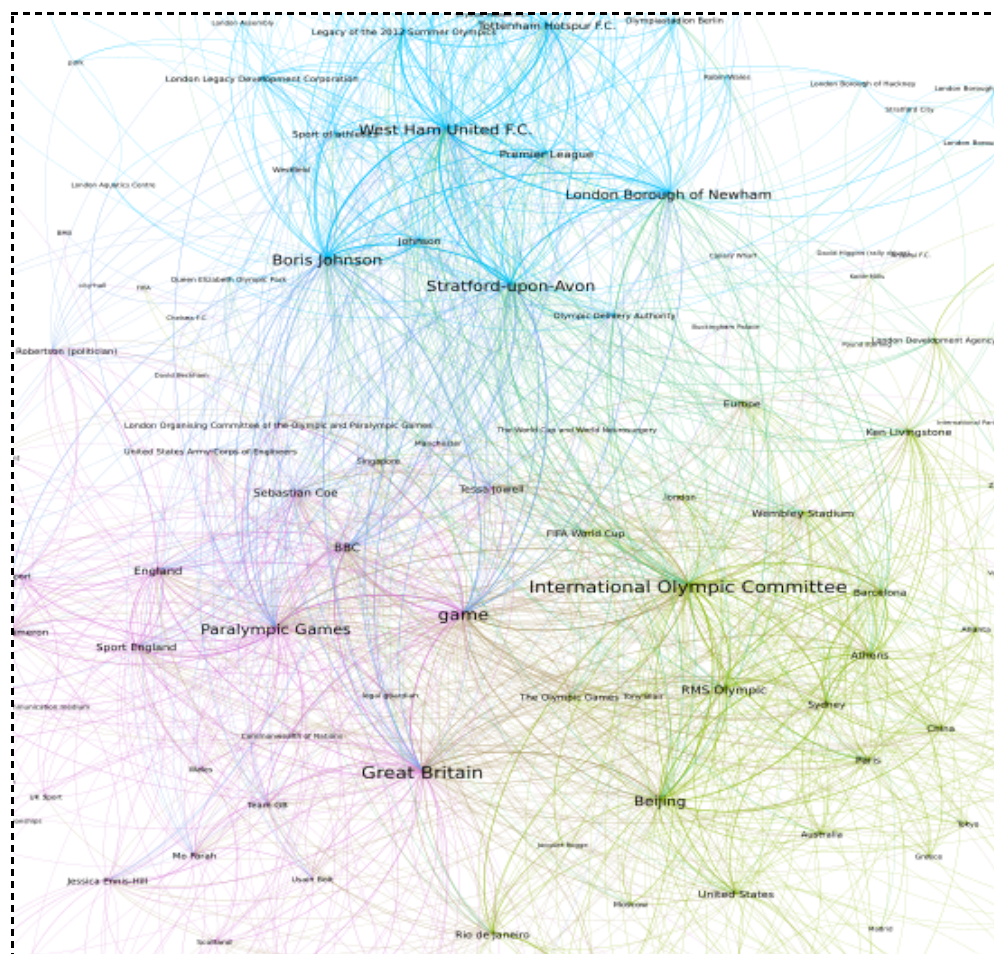
The network presented above is composed of three clusters (blue, green and purple). Central entities in the Blue cluster are Stratford-upon-Avon[41], Boris Johnson, West Ham United F.C., and the London Borough of Newham. In Purple, it is possible to highlight the Paralympic

---

[41] This is an error generated by linking Named Entities with Wikidata. The correct entity should be Stratford (east London).

Games, Great Britain, game, and BBC; in Green, International Olympic Committee and Beijing. A more detailed analysis of the network is presented in the following chapters.

3.6 Natural language processing of textual information

Natural language processing or simply NLP is a set of techniques and a sub-area of Artificial Intelligence (AI) frequently used by researchers in what is called 'computational humanities' (Kumar, 2011). Instead of using Digital Humanities as an umbrella term, it is important to emphasise here the computational aspect, as NLP is situated between DH and computer science. NLP techniques have been developed to analyse and process natural language data. They use trained models to capture the structure of the language and its particularities, allowing the user to find patterns, the co-occurrence of words and connections between different terms, as well as identifying sentiments, entities, events and contexts that appear in a text. In this research, several approaches to NLP were used, most of them based on Python libraries such as Spacy (Honnibal and Montani, 2017), to search for data and scrape web pages. The work of data harvesting and preparation involved saving data from the web, extracting the information from documents, transforming text into usable data and the visualisation of that data in order to assist with analysis.

Many times, the process of building these tools is considered a 'black box' (Mannarswamy and Chidambaram, 2021). It can be difficult for a digital humanist to determine how the algorithms were designed and what tasks they actually perform. When using these sorts of methods for DH, many questions emerge, for example what exactly does a model consider as positive or negative when analysing sentiment in a text? The nuances and particularities of a language can also be lost due to the use of quantitative - generalising - methods, even more so when considering under-resourced languages - those with a lack or a small number of trained models. Because of this, the methodology for this research was based on successive testing of multiple techniques in order to find the best or most appropriate combination of different techniques.

To have an overview of the data based on the most common words used by journalists to cover the Olympic legacies, the Natural Language Toolkit – NLTK (Bird, Kein and Loper, 2009) was used. It is a set of Python methods developed to manage text analysis. By using, for instance, its tokenizer tool it was possible to write a ranking of top mentioned words in the corpus into a spreadsheet. Moreover, it allowed for the identification of bigrams[42] - the most used combinations of two tokens (words) - as well as TF-IDF[43] - a statistical formula to determine the most important words in a document. The following spreadsheet shows what the output looked like. We have for each column a different news outlet – or a combination of outlets – and each line has a word followed by its frequency (see Appendix F and G).

Figure 14 – Spreadsheet with word trends extracted from British sources using the NLTK

	A	B	C	D
1	**Top words all**	**Top nbigram all**	**Top words tfidf**	**top news**
2	('games', 8386)	('west ham', 1317)	('games', 0.39959768623194253)	('games', 7264)
3	('sport', 4977)	('world cup', 556)	('sport', 0.23715689057672049)	('sport', 4955)
4	('new', 4223)	('long term', 511)	('new', 0.2012283602381938)	('people', 3890)
5	('people', 4215)	('young people', 487)	('people', 0.2008471556722678)	('stadium', 3816)
6	('stadium', 3666)	('boris johnson', 433)	('stadium', 0.1746869923355952)	('new', 3314)
7	('park', 3356)	('chief executive', 427)	('park', 0.1599153154059622)	('sports', 2577)
8	('city', 3122)	('uk sport', 417)	('city', 0.14876508185262635)	('park', 2457)
9	('world', 2872)	('paralympic games', 390)	('world', 0.13685243916743847)	('year', 2440)
10	('sports', 2566)	('lord coe', 368)	('sports', 0.12227136452076848)	('world', 2297)
11	('year', 2496)	('lea valley', 365)	('year', 0.11893582456891587)	('time', 2109)
12	('time', 2272)	('sport england', 356)	('time', 0.10826209672298753)	('city', 1960)

Figure 14 shows four columns that are part of the spreadsheet. Column A shows the most frequently mentioned words in the entire dataset; Column B shows the most mentioned bigrams in the entire dataset; Column C shows the most mentioned words using TF-IDF;

---

[42] An n-gram is a sequence of n words. A bigram is a sequence of two words.

[43] Term Frequency – Inverse Document Frequency (TF-IDF) is a technique used to count words in a document by computing weights for each word according to their importance in the corpus.

Column D shows the most mentioned words in the subset 'news', composed solely of articles published by news outlets. Other columns in the spreadsheet can be seen in the Appendix F.

The spreadsheet is helpful for viewing detailed information but it may not be the clearest way to represent this set of trending topics. For more effective visualisation, word clouds, as shown in the example below (Figure 15), were plotted. Word clouds also have their limitations. Word clouds represent a set of words mentioned in a document. The more a specific word appears in the document, the bigger and bolder it appears in the cloud. As the human brain cannot easily compare word sizes and hues (Cairo, 2012), this figure does not allow detailed comparisons, such as how much bigger 'year' is in relation to 'people'. However, it is an efficient mechanism to represent huge contrasts like the sizes of the words 'stadium' and 'nation'. Among the most prominent words in the Figure are 'stadium', 'year', 'people', 'park', 'sports', 'city', 'world', 'team', 'public', 'last', 'home', 'government', 'housing', 'venus', 'bid', 'British', 'development', among others.

Figure 15 – Word cloud showing the most frequently occurring words in the entire dataset composed of corpora covering London Olympics in English

It is worth looking at the ranking of bigram collocations extracted from the texts using the NLTK. This combination of two words that are recurrent in the data is an effective method to have an overview of the most discussed topics. For the purpose of this research, bigrams are

more useful than single words (tokens) as the combination often helps to qualify the words or at least provide us with clues to specify their connotation – which must be further analysed in depth by close reading of a selection of the articles. In the tables below, the top most frequently occurring bigrams in the news articles covering London's and Rio's Olympic legacies are listed. The articles were divided into three categories: news, government and activist. The first one refers to the set of articles published by the news media, the main focus of this study. The second is a list from articles published by official government websites such as uksport.gov.uk. The third category is composed of activist blogs created to protest against the Olympics or raising concerns about the Olympic legacies. Although this thesis is mainly focused on the first category, the others were collected to help with identifying the unsaid. Looking at government official websites as well as activist blogs revealed topics that were not discussed in the news media, and therefore, represent 'silences' in the journalistic coverage of those events.

Initially, 50 bigrams were generated using the NLTK. After cleaning – removing stop-words[44] – the words below were organised in columns followed by a number that represents how many times each bigram was found:

Table 1 – Top bigrams mentioned in articles published in English about London's legacy

top bigram news	top bigram gov	top bigram activist
('west ham', 1446)	('international inspiration', 119)	('stratford city', 254)
('sport england', 475)	('world class', 108)	('mega event', 212)
('world cup', 471)	('world cup', 97)	('lea valley', 202)

---

[44] Even though many stop-words were added to the list for deletion from the final spreadsheet, some irrelevant words such as 'click here' were kept as they are easily identifiable.

('young people', 457)	('young people', 97)	('clays lane', 180)
('boris johnson', 447)	('queen elizabeth', 93)	('new park', 167)
('long term', 436)	('national lottery', 93)	('mega events', 160)
('chief executive', 421)	('elizabeth park', 92)	('julian cheyne', 153)
('lord coe', 402)	('paralympic games', 80)	('new jobs', 132)
('school sport', 305)	('world championships', 63)	('long term', 127)
('prime minister', 300)	('development corporation', 61)	('public housing', 122)
('international committee', 297)	('major events', 59)	('local people', 121)
('tessa jowell', 265)	('lee valley', 54)	('delivery authority', 120)
('opening ceremony', 256)	('chief executive', 51)	('lower lea', 120)
('premier league', 252)	('gold event', 47)	('leabank square', 120)
('team gb', 238)	('event series', 47)	('open space', 115)
('paralympic games', 230)	('sporting events', 40)	('paralympic games', 113)
('uk sport', 216)	('great britain', 37)	('british cycling', 112)
('gold medal', 210)	('children young', 36)	('europe 150', 108)
('aquatics centre', 207)	('east bank', 34)	('urban park', 96)

('new homes', 206)	('high performance', 31)	('real estate', 94)
('ken livingstone', 198)	('aquatics centre', 29)	('jobs created', 94)
('local people', 197)	('england hockey', 29)	('arena fields', 92)
('delivery authority', 194)	('european championships', 28)	('hackney wick', 91)
('organising committee', 194)	('look forward', 27)	('health safety', 90)
('sports minister', 193)	('inspiration programme', 27)	('park europe', 90)
('east end', 190)	('lottery funding', 26)	('compulsory purchase', 86)
('park company', 189)	('team gb', 24)	('affordable housing', 85)
('physical activity', 181)	('elite sport', 23)	('waltham forest', 84)
('development corporation', 179)	('million children', 23)	('local community', 83)
('ham united', 179)	('eurohockey championships', 22)	('athletes village', 82)
('lea valley', 178)	('british athletes', 22)	('public funds', 80)
('playing fields', 176)	('women world', 21)	('local residents', 75)
('football club', 173)	('long term', 21)	('urban renewal', 72)
('david cameron', 172)	('championships 2015', 21)	('construction sector', 72)
('mayor boris', 171)	('international development', 21)	('result games', 72)

('governing bodies', 169)	('people world', 21)	('related mega', 72)
('public money', 168)	('class events', 20)	('mountain bike', 72)
('development agency', 166)	('commonwealth games', 19)	('post games', 68)
('queen elizabeth', 165)	('local people', 18)	('urban governments', 66)
('elizabeth park', 163)	('loughborough university', 18)	('media centre', 66)
('school sports', 159)	('valley hockey', 18)	('largest new', 66)
('million people', 157)	('copper box', 17)	('martin slavin', 65)
('world class', 157)	('box arena', 17)	('sports facilities', 63)
('athletes village', 146)	('inspire generation', 17)	('international committee', 60)

Table 2 – Top bigrams mentioned in articles published in Portuguese about Rio's legacy

top bigram news	top bigram gov	top bigram activist
('jogos olímpicos', 438)	('ministério esporte', 129)	('vila autódromo', 174)
('ministério esporte', 274)	('jogos olímpicos', 92)	('jogos olímpicos', 133)
('parque barra', 231)	('parque barra', 67)	('click here', 39)
('autoridade governança',	('alto rendimento', 58)	('here english', 39)

198)		
('barra tijuca', 187)	('confederação brasileira', 37)	('pagina inicial', 38)
('governo federal', 180)	('arena carioca', 35)	('barra tijuca', 35)
('eduardo paes', 161)	('autoridade governança', 34)	('maria penha', 24)
('ano passado', 154)	('últimas notícias', 32)	('theresa williamson', 22)
('centro tênis', 149)	('centro tênis', 32)	('eduardo paes', 20)
('governança aglo', 145)	('ministro esporte', 31)	('zona norte', 20)
('arena carioca', 125)	('nacional esporte', 31)	('durante jogos', 20)
('ministério público', 112)	('controle dopagem', 27)	('copa mundo', 19)
('zona oeste', 106)	('rede nacional', 26)	('plano popular', 19)
('parque radical', 105)	('gov br', 26)	('moradores vila', 18)
('alto rendimento', 100)	('leonardo picciani', 26)	('zona oeste', 18)
('iniciativa privada', 88)	('governo federal', 25)	('direitos humanos', 18)
('confederação brasileira', 85)	('vôlei praia', 25)	('últimos anos', 17)
('público federal', 84)	('nacional treinamento', 24)	('milhões árvores', 17)
('prefeito eduardo', 83)	('esporte alto', 24)	('los angeles', 17)

('arena futuro', 81)	('governança aglo', 24)	('comunidades catalisadoras', 16)
('rock in', 81)	('canoagem slalom', 22)	('moradores favelas', 16)
('baía guanabara', 80)	('centro treinamento', 21)	('museu remoções', 15)
('prefeitura janeiro', 79)	('paulo márcio', 21)	('sediar jogos', 15)
('instalações esportivas', 74)	('durante jogos', 20)	('baixa renda', 14)
('centro treinamento', 74)	('francisco medeiros', 20)	('porto maravilha', 13)
('instalações olímpicas', 72)	('universidade federal', 17)	('especulação imobiliária', 13)
('comitê organizador', 72)	('pan americano', 17)	('prefeito eduardo', 13)
('comitê internacional', 71)	('tênis mesa', 17)	('comitê internacional', 13)
('leonardo picciani', 68)	('foto francisco', 17)	('políticas públicas', 12)
('corpo bombeiros', 67)	('jiu jitsu', 17)	('planejamento urbano', 12)
('bbc brasil', 66)	('notícias tweet', 16)	('https co', 12)
('zona sul', 65)	('tiro esportivo', 16)	('feira jul', 12)
('radical deodoro', 65)	('ricardo leyser', 15)	('zona sul', 12)
('canoagem slalom', 62)	('paralímpico brasileiro', 15)	('mídia internacional', 12)
('ministro esporte', 58)	('esporte brasileiro', 15)	('linhas ônibus', 12)

('arenas cariocas', 58)	('esporte leonardo', 15)	('cidades sede', 12)
('velódromo parque', 57)	('circuito mundial', 15)	('dona penha', 12)
('marcelo crivella', 56)	('atletas brasileiros', 14)	('andrew zimbalist', 12)
('olímpicos paralímpicos', 56)	('atletas alto', 14)	('ano olimpíadas', 11)
('paulo márcio', 56)	('olímpicos paralímpicos', 14)	('baía guanabara', 11)
('cidade janeiro', 55)	('rededoesporte gov', 14)	('inicial olhonasremoções', 10)
('comitê brasil', 53)	('secretaria especial', 13)	('complexo alemão', 10)
('velódromo centro', 52)	('george hilton', 13)	('região portuária', 10)
('esporte sediou', 52)	('bolsa atleta', 13)	('cidade janeiro', 10)
('ministério cidadania', 51)	('lei incentivo', 13)	('longo prazo', 10)
('campo golfe', 51)	('comitê paralímpico', 13)	('ano jogos', 10)
('tribunal contas', 50)	('desenvolvimento esporte', 13)	('24 milhões', 10)
('autoridade pública', 50)	('preparação atletas', 13)	('nolympics la', 10)
('zona norte', 49)	('comitê brasil', 13)	('brt transcarioca', 10)
('linha metrô', 49)	('centros treinamento', 13)	('heloisa helena', 10)

By looking at these bigrams and other outputs for word trends generated from the corpus, I seek to identify the main topics of interest in the media coverage.

3.6.1 TIME: a tool for temporal analysis of news coverage

These word trends were used to produce a demonstrator as one of the results of CLEOPATRA's Research & Development weeks. The tool called TIME[45] produces charts for a temporal reading of the trends, where it is possible to analyse the timeframe in which one word was most mentioned in comparison to others. The demo works entirely in English as it uses a translated version of the Brazilian data. TIME is based on the curated sub-dataset mentioned above in this chapter. The tool was created in collaboration with CLEOPATRA early-stage researchers Daniela Major (School of Advanced Study), Gullal Cheema (Leibniz Information Centre for Science and Technology) and Abdul Sittar (Jožef Stefan Institute). The chart below (Figure 16) illustrates the functionalities of TIME to visualise word trends over time. At the top of the page, the user can add as an input the media outlet ('bbc', for example) and the word they are interested in analysing ('stadium', for example). The chart shows how many times the word was mentioned by the BBC over the years.

---

[45] TIME is available at *http://cleopatra.ijs.si/dashboard/.*

Figure 16 – Chart provided by TIME to visualise word trends over time of articles covering London 2012 and Rio 2016 Olympic legacies
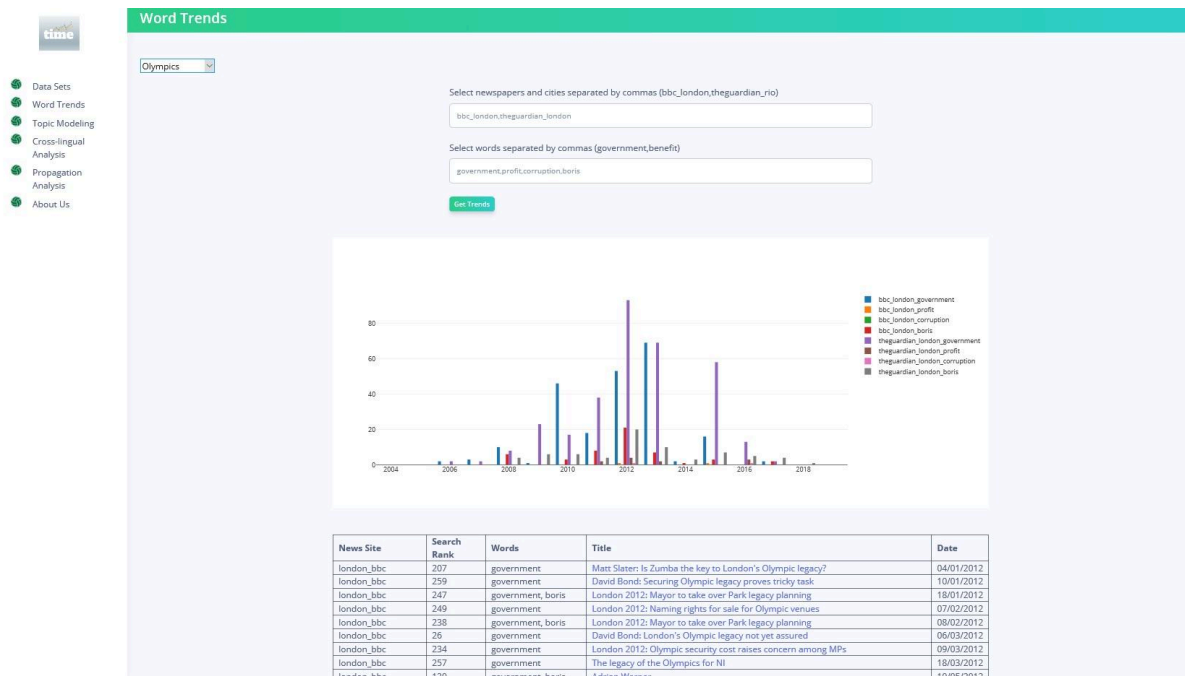
Besides providing charts for temporal analysis, by clicking on one of the bars the user can visualise the article's title, date and position in the Google search rank. By clicking on one of the titles, the user is shown the full article text, with the linguistic entities mentioned in the text highlighted and labelled, as shown in the figure below. Here, the number 'four' was labelled as 'cardinal'; 'Radio 4' as 'organisation'; 'You and Yours' as 'work of art'; 'London' as 'geopolitical entity'; 'BBC' as 'organisation', and so on.

Figure 17 – Article body text with entity recognition

Another affordance of TIME is topic modelling. TIME provides charts to explore the topics covered by the media using topic modelling algorithms. Topic modelling is a machine learning technique capable of identifying patterns in a set of words present in a document, clustering the words in groups and finding expressions that are characteristic of this document (Graham, Weingart and Milligan, 2012). Figure 18 shows some outputs using topic modelling in TIME.

Figure 18 – Topic modelling performed for texts published by the British media covering the London 2012 Olympic legacy
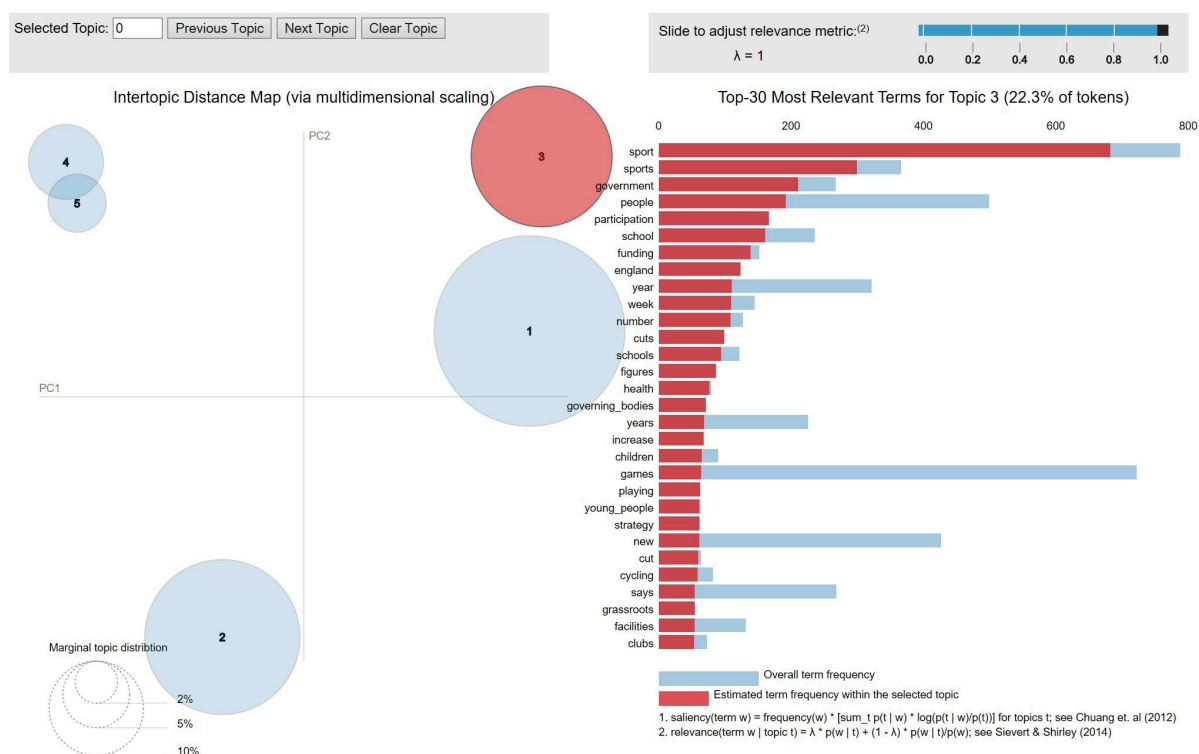
The circles on the left-hand side show the generated topics and how close or distant they are from each other. Topics four and five, for example, share relevant terms given their proximity. On the right-hand side, the chart shows the top-30 most relevant terms for each topic (in this figure, data is shown for topic three). In red are represented how many times each word appears within the topic. The blue bars, for comparison, show how many times the same word is mentioned in the entire dataset.

It is important to mention, however, some limitations imposed on the functionality and use of this tool. TIME is hosted in the Jožef Stefan Institute's server in Slovenia, one of the institutional partners in the CLEOPATRA project. Sometimes, if there is an error in the server, or if it is for any reason shutdown and restarted, the tool becomes unavailable and technical staff based in Slovenia need to be contacted via e-mail and perform the adjustments when they have available time. Because of this, the tool was regularly inoperative during this

research project. When available, TIME was used due to the affordances outlined above and its usability. On the other hand, when not working, other tools such as Voyant Tools or python libraries such as Spacy were used for textual analysis.

These NLP techniques described above will be referred to in the next chapters as key methods used for the data analysis.

3.7 Positive and negative legacies: sentiment analysis

The following headlines illustrate how negative legacies are framed in the media: 'Rio's Olympic legacy a "huge disappointment"' (Davies, BBC, 2017); 'Brazil's $12 billion Olympic legacy lies in ruins' (Daily Mail, 2017); 'Key London Olympic legacy "a failure", says Tessa Jowell' (BBC, 2015). By googling 'London's Olympic legacy' or 'Rio's Olympic legacy', it is not difficult to find news articles whose titles indicate negative aspects of the games. However, there seems to be a significant difference between both searches. It was very hard to find something positive about Rio's legacy to use as an example here. When looking for London's legacy positive results do come up in the search. These 'positive' headlines usually cover topics such as sustainability and the regeneration of East London. In particular, if searching in Portuguese for *legado olímpico de Londres*, the disparity seems to be even more relevant, with significantly more positive articles, such as 'Six years later, the wonderful legacy of London 2012' (Globo, 2018) and 'London and its Olympic legacy: East-side of London still growing 5 years after the games' (Globo, 2017).

Conducting an online key-word search is not sufficient to get an understanding of positive and negative sentiments, and analysing the texts one by one is a laborious task. There are methods of NLP and computational linguistics developed to capture the sentiment expressed in a text. These methods are called Sentiment Analysis (SA). In this research, I explore the benefits of using SA to analyse the sentiments in huge amounts of texts. While it is possible to infer some hypotheses about the use of positive and negative words around the topic of

'legacy' through close reading the articles found via Google search, SA techniques can be useful for providing an overview over larger amounts of data.

I will test the suitability of different techniques and models for opinion mining in news articles. While most of the models for SA target publications on Twitter or e-commerce, I will focus on the use of SA to track sentiments expressed in journalistic texts covering a specific topic: the legacy of the last two Olympics. The main objective in applying this method is to make use of distant reading approaches to reach a broader comprehension of the positive and negative sentiments expressed in the media coverage of the games.

Two different concepts have been merged here and require special attention. When using the terms positive and negative legacies, we refer to the dimensions of legacy stated by Gratton and Preuss (2008) and approached in this thesis above. Positive and negative sentiments, by contrast, are classifications used by SA to categorise the emotions expressed in a text. However, I hypothesise that there is a correlation between the use of negative terms – therefore a negative sentiment – and the reporting of negative legacies.

3.7.1 Sentiment analysis as a method for the humanities

Sentiment Analysis (SA) is often defined as a method for identifying opinions about specific entities in a text. It is frequently used to track the reputation of companies and products, especially on social networks like Facebook and Twitter (Feldman, 2013). By using SA, companies can monitor their brands and get feedback in real time about consumers' opinions. Most of the challenges involved in SA approaches pointed out by researchers arise from the specificities of the kinds of communication that are present on Twitter and Facebook. Sarcasm, problems with spelling and grammatical mistakes are often cited as barriers for the development of good algorithms (Maynard, 2014). However, these are not the main problems faced when taking news articles as objects of study. An issue that does apply to this corpus is the accuracy of approaches that tend to target subjective statements and overlook objective

ones. As news articles and reportage tend to use objective language, traditional approaches of SA need to be tested and analysed accordingly (Balahur & Steinberger, 2009).

Understanding the origins and traditional applications of SA leads us to a critical reflection on its usability for humanities research. This process of recognising problems with the data and the applied methods is inherent to the usage of digital methods (Rogers, 2019, p. 3). As it is heavily used in the context of e-commerce with a focus on reviews of products and services, SA algorithms tend to look at words, mainly adjectives, often used by customers. When we change the objective to track sentiment about media events using news articles, the models used provide results that require further analysis to be interpreted according to the context in which they have been applied. Another issue to be considered is the uncertainty involved in using models trained in different contexts and on different media and text types for this very specific corpus, or what is called domain-specificity.

Furthermore, language is a widely-recognised barrier for SA (Dashtipour et. al. 2016; Reis et al. 2015; Saldaña 2018). Due to the lack of efficient models trained in Portuguese, it is necessary to use translation techniques, as the most common approach to SA in languages other than English. Although questions are raised about the loss of meanings through the use of machine translation in SA, the study of SA for texts written in Portuguese has shown a significant level of accuracy, as long as correctly validated (Reis et al. 2015).

As suggested by Gonçalves et al. (2013), this study combined different widely used SA algorithms, in order to reach a higher level of accuracy. Among the algorithms used were SentiStrength (Thelwall et. al., 2010), SenticNet (Cambria and Hussain, 2015) and Vader (Hutto and Gilbert, 2014). These lexicon based algorithms split the text into words and look for their scores in the dictionary (Thelwall, 2014). They cannot disambiguate meanings as they do not always take the context into account. Two contextual based algorithms were also used for this corpus to get scores for texts in English: Amazon BERT[46](McAuley and

---

[46] BERT: Bidirectional Encoder Representations from Transformers is a machine learning technique for NLP developed by Google.

Leskovec, 2013; Davlin, 2019) and Sent140 BERT (Go et al., 2009). The body text score is the average of the scores attributed to each sentence in the article.

For texts and news headlines in Portuguese, SenticNet, SentiStrength and Portuguese Twitter BERT[47] were applied. Besides using these three algorithms, texts and titles were translated into English and the algorithms mentioned above (for English classification) were also used for the translated version of articles originally published in Portuguese. The main objective of translating and applying the same algorithms used for English was first, to allow comparisons between Portuguese and English scores and second, to evaluate both methods in order to find the best combination of techniques for journalistic data exploration using SA.

The main reason for using SA in this research is to first understand how positive and how negative the media coverage of the Olympic legacies of the London 2012 and Rio 2016 Olympic Games was. SA was also used to track how much 'affective language' was used by the newspapers to report on these events. By affective language, I mean the use of adjectives, adverbs or any qualifiers that intensify the words used in the text.

In total, eight algorithms were applied to the corpus. For articles published originally in English or translated into English, five algorithms were applied. The following table summarises the distribution of algorithms for each of the two languages used:

Table 3 – Summary of sentiment classifiers applied to the dataset for each language

	SentiStrength	SenticNet	Vader	Amazon BERT	Sent140 BERT	Portuguese Twitter BERT
English	x	x	x	x	x	

---

[47] Documentation available at
https://www.kaggle.com/augustop/portuguese-tweets-for-sentiment-analysis.

Portuguese	x	x				x

A final result was produced based on the 'best of three' rule. If three algorithms agreed on one sentiment label – positive, neutral or negative – the final result was determined by this agreement. Where they did not agree in at least three columns, the result was labelled as 'uncertain' and was used later on to analyse possible reasons for this disagreement. The best of three was used as a proxy for reviewing and verifying the algorithmic labels. It corresponds to getting three humans to label the data in order to ensure there is consensus between annotators. In the case of humans, they would ideally be domain experts. SA algorithms however must be read taking into consideration their potentials and limitations, which are further discussed later in section 3.7.3 of this chapter.

3.7.2 Data preparation for SA

A few preprocessing steps were taken before the texts were ready for SA. Spacy (Honnibal and Montani, 2017), a natural language processing library, was used to segment the whole article into sentences with language-specific models for both English and Portuguese. Previous work on multilingual (Dashtipour, 2016; Zhang, 2018) and Portuguese text sentiment analysis (Cirqueria, 2016; Farias, 2016; Pereira, 2021; Tavares, 2021) has often used translation-based methods as baselines to compare against language specific approaches. As demonstrated by Araújo et al. (2016, 2020) and Cirqueria et al. (2016), machine translation from Portuguese into English significantly improved the sentiment results in comparison to the use of non-English classifiers. Having said that, previous work also refers to mistakes introduced into SA methods after translation into the English language. To investigate the impact of translation, for news articles written in Portuguese language, I used the google-translate API (Han, 2020) for translating text into English and compared the English language models' performance on translated text with classifiers trained with and for Portuguese language.

One further issue identified when preparing the data was that the words 'Olympic' and 'legacy' appear in the entire corpus and seemed to carry positive sentiment, based on preliminary experiments. As a result, an alternative sub-dataset was prepared by replacing different instances of 'Olympic' with 'event' and 'legacy' with 'outcome' to avoid possible bias. This word replacement was performed in both English and Portuguese texts, a summary of which is provided in Table 4.

Table 4 – List of words replaced to avoid bias in the corpus

Words	Replacement
Olimpíadas, olimpíadas, Olimpíada, olimpíada	evento
Olímpicos, olímpicos, Olímpico, olímpico, Olímpicas, olímpicas, Olímpica, olímpica	do evento
Legado, legado, Legados, legados, heranças, herança, Heranças, Herança	resultado
Olympic, Olympics, olympics, olympic	event
legacy, Legacy	outcome

3.7.3 Evaluating the SA classifiers

The classifiers detailed above were used to assign sentiment labels to news headlines and whole articles. For assigning a sentiment label to the whole article, the choice was made to use the proportion of sentences of each sentiment type instead of averaging the sentence scores. With averaging scores, there is a chance of negative and positive scores cancelling each other out and the final score being closer to zero (neutral). The proportion of sentences with a particular sentiment type provided a much clearer view of the narrative used in the article. For example, if the proportion of sentences with negative labels is greater than 50%, then the article is assigned a negative label. Figure 19 shows the distribution of positive, neutral and negative sentiment labels produced by each classifier for data in English and Portuguese.

Figure 19 – Distribution of models' outputs (positive, neutral and negative per classifier for data originally in English, data originally in Portuguese and data translated from Portuguese into English. Data in Portuguese contain 464 articles and data originally in English contain 807. PT/EN means Portuguese text translated to English
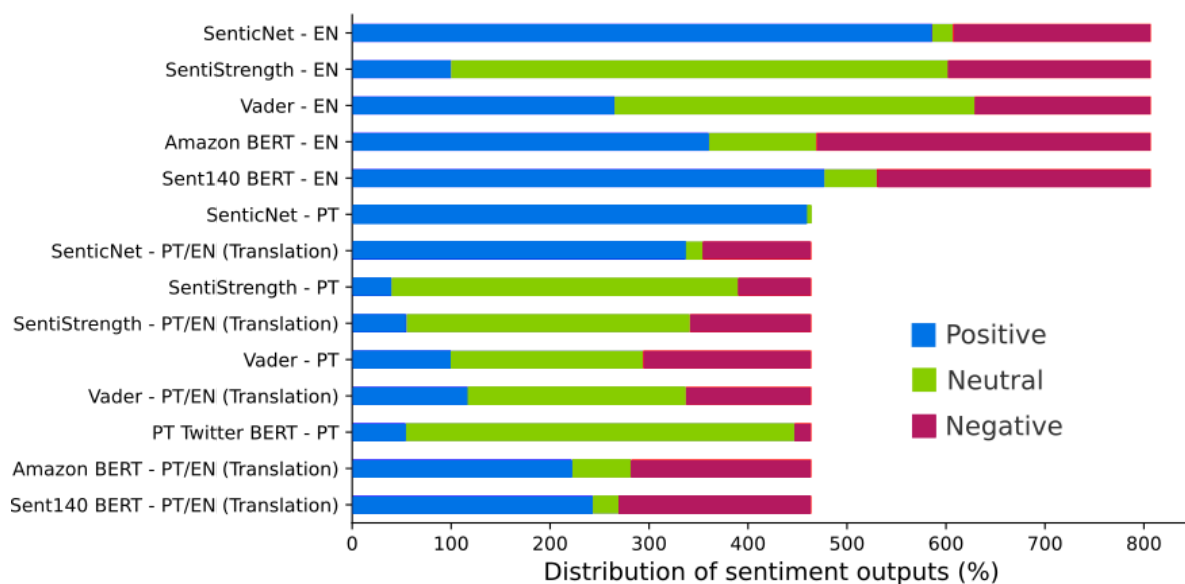
To evaluate the performance of the algorithms, both the sentiment expressed by the news articles and the news headlines of four outlets – two in English and two in Portuguese – were labelled by the author of this thesis, who has expert knowledge on the topic. As a domain expert, I assigned positive, neutral and negative labels to 717 news articles (56.4% of the total number of articles).[48] This list of what I called 'gold labels' was used as a reference point to compare with the results generated by each classifier, in order to evaluate their accuracy for this specific dataset.

The choice of using news headlines was made for a number of reasons. First, they are shorter to read and annotated by a human but also less complex to be classified automatically by the algorithms than the entire article. The challenge with classifying a whole article is not only the length of the text but also the complexity of the multiple arguments usually presented within a news article – many times privileging contradictory perspectives. These characteristics can be seen in *'Olympic legacy: White elephant or economic viagra?'* (Anderson, BBC, 2012). The title already informs us that this text will cover both positive (fast economic growth) and negative (unused infrastructure) aspects of the legacy. The text starts by reporting on the success of the London Olympics, referred to as 'unequivocally the most successful Olympic Games for Britain':

> measuring the sporting success of London 2012 is easy - a quick glance at the medals table tells us unequivocally that this has been the most successful Olympic games for Britain in more than a century (Anderson, BBC, 2012).

However, the next section of the article, entitled 'scared off', presents a negative perspective: shops did not see the expected increase in the number of customers during the Games. The last section, entitled 'Fabulous city', although very positive, also includes sentences such as 'there is a feeling in Sydney and Athens that they have failed to fully capitalise on the

---

[48] The dataset is published and available at
https://github.com/caiocmello/sentiment-annotation-olympic-news.

Games', where the idea of failure is stated, even though not referring to the London Olympics. This illustrates the complexity of analysing the entire body text of news articles.

Moreover, news headlines are the first and more important piece of information in the news articles, being sometimes the only part that readers have access to, when we consider news websites under paywalls or how news outlets spread their content on social media. Content on Instagram or Twitter, for example, is often composed of an image, the news headline and a URL to access the website.

The assumption that news headlines intend to summarise the article's content in one sentence is however challenged by the literature. Scacco & Muddiman (2016) divided news headlines into three categories: (1) traditional news headlines, those following the inverted pyramid-structured. This means they prioritise the most important information, and therefore could be taken as a good summary of the content. However, there are other kinds of titles, such as (2) forward-reference headlines, which emphasise unknown information. In this case, headlines hide the main content in order to guide the reader to read the entire article. And lastly there are (3) question-based headlines, which are focused on uncertainty. By stating a question, they indicate that the answer can be obtained by reading the full text.

Considering that news headlines are not necessarily a summary of the content but have a significant importance considering the forms in which news are produced and distributed,[49] they are taken as objects for this study's analytical process.

The following table shows the distribution of sentiment outputs across the dataset.

---

[49] Headlines are particularly important to call the attention of readers to either buy a newspaper or to click on a link to the news webpage. They are key components to convey quick and short messages but also essential players in the commercialisation of the news: the more interesting the headlines are, the more visits a news website will have. This aspect of news headlines is further discussed in Chapter 7.1.

Table 5 - Distribution of positive, neutral and negative sentiment outputs for each cluster used to analyse classifiers' accuracy, followed by *gold labels* distribution

	News outlet/Event											
	Globo (Rio)			Estadao (Rio)			Guardian (London)			Daily Mail (London)		
	Pos	Neu	Neg	Pos	Neu	Neg	Pos	Neu	Neg	Pos	Neu	Neg
PT SenticNet	243	2	0	119	1	0	x	x	x	x	x	x
PT SentiStrength	20	184	41	15	89	16	x	x	x	x	x	x
PT Vader	57	103	85	25	49	46	x	x	x	x	x	x
PT Twitter BERT	7	234	4	13	104	3	x	x	x	x	x	x
EN SenticNet	179	8	58	92	3	25	167	1	49	90	4	41
EN SentiStrength	26	151	68	19	69	32	23	149	45	24	67	44
EN Vader	69	110	66	26	56	38	61	107	49	44	56	35
EN Amazon BERT	140	21	84	45	21	54	94	33	90	52	16	67
EN Sent140 BERT	135	11	99	57	8	55	144	11	62	65	12	58
Gold Labels	79	64	102	32	28	60	93	55	69	44	26	65
Total Titles	245			120			217			135		

The scores, converted into three labels (positive, neutral, negative), were compared with the gold labels. The results of this comparison for news headlines are shown in Table 6.

Table 6 – (%) Matches of each classifier with the gold labels (accuracy evaluation for this dataset) for news headlines. PT: Portuguese; EN: English. Data distribution (number of articles): Globo (Rio): 245; Estadao (Rio): 120; Guardian (London): 217; Daily Mail (London): 135.

Cluster (Outlet/Event)	PT Sentic Net	PT SentiStrength	PT Vader	PT Twitter BERT	EN Sentic Net	EN SentiStrength	EN Vader	EN Amazon BERT	EN Sent140 BERT
Globo (Rio)	33%	39.5%	49.7%	24.5%	46.5%	50.2%	55.5%	55.5%	61.2%
Estadao (Rio)	26.4%	15.7%	39.6%	3.3%	38.8%	31.4%	40.4%	53.7%	57%
Guardian (London)	x	x	x	x	48%	49.6%	56.2%	60%	58.5%
Daily Mail (London)	x	x	x	x	52%	47.9%	56.6%	61.7%	61.2%

The table shows how many matches were found when analysing news headlines and comparing the labels produced by each of the sentiment classifiers with the gold labels assigned by me. As expected,[50] the English translation performed much better than the original text in Portuguese. Among all classifiers, the best results were obtained from Amazon BERT and Sent140 BERT - around 61% for Daily Mail news titles covering the London 2012 Olympics.

To obtain the best accuracy, I opted for combining the result of the three algorithms with best performance as stated above: Vader, Amazon BERT and Sent140 BERT. This way, if at least two of them agreed, having assigned the same label to the sentence, this answer was taken as

---

[50] Araújo et al. (2020); Pereira (2021) and De Freitas & Vieira (2015) discussed the limitations of working with data in Portuguese.

likely correct. Otherwise, if each algorithm gave a different result, the sentence was considered 'inconclusive' and it was saved in a separate folder for further analysis.

3.7.4 Potentials and limitations of the SA

As pointed out by Rogers (2019, p. 3), the process of recognising problems with the method and the data is inherent to working with digital methods. Working with this methodology of combining multiple classifiers culminated in some inconsistencies that are discussed in this section.

First, it is important to point out the disagreements between the expert and the machine. For the headlines 'Martin Samuel: Tessa Jowell deserves an Olympic medal in utter madness' (Samuel, Daily Mail, 2010), 'Olympic legacy: school sports provision patchy across UK, admits Jeremy Hunt' (The Guardian, 2012) and 'Letters: the true Olympic legacy is white elephants on our doorstep' (Primm, 2012) the tool labelled them as positive utterances while I labelled them as negative. It is hard to explain why the machine assigned some labels to specific sentences, but there are mechanisms for hypothesising the cause.

One of the reasons why models make such mistakes can be attributed to the domain gap, which is the difference between the domain of training data – such as primarily social media posts – versus the evaluation data – news headlines and text, for example. Cross-domain sentiment analysis is a challenging and fairly active research area (Al-Moslmi, 2017). While many techniques exist to mitigate the problem, most require a variant of cross-domain learning using vast source data, with labels, and target data – such as news. The objective of the learning process is to get a better performing sentiment detection model for the target data. Training models on such data was, however, not viable within the scope of this thesis as it is costly and time-consuming. Therefore, in this study, I rely on the existing publicly available monolingual sentiment detection models and investigate the model outputs with a fine-grained analysis in order to account for the limitations of these models.

To hypothesise on how the algorithms work, I used explainable AI, a mechanism to unveil the machines' behaviour (Linardatos et al., 2021), performed in collaboration with a colleague from the CLEOPATRA project, Gaurish Takkar. We used a tool named SHAP (Lundberg & Lee, 2017) on the two best performing classifiers: Amazon BERT and Sent140 BERT. The tool used Shapley values to depict how fairly the outcome can be distributed among the different features. Thus, the Shapley values of every feature signal its contribution for a given predicted (into negative, positive or neutral) instance. We utilised an existing tool to quantify the feature importances for BERT-based models. The input to the tool is a text and the classifier. Shap returns a pictorial depiction of Shapley values for every feature (Figure 20), which are words in this case. The output enables easy examination of the predictions. Figure 20 illustrates SHAP visual representation for the news headline 'Olympic legacy: school sports provision patchy across UK, admits Jeremy Hunt' (The Guardian, 2012).

Figure 20 – SHAP depiction of Shapley values. Only the words considered as having significant contribution to the final classification are shown. The negative sentiments are represented in red. The positive, in blue. The label assigned to this sentence by the combination of classifiers was *positive*
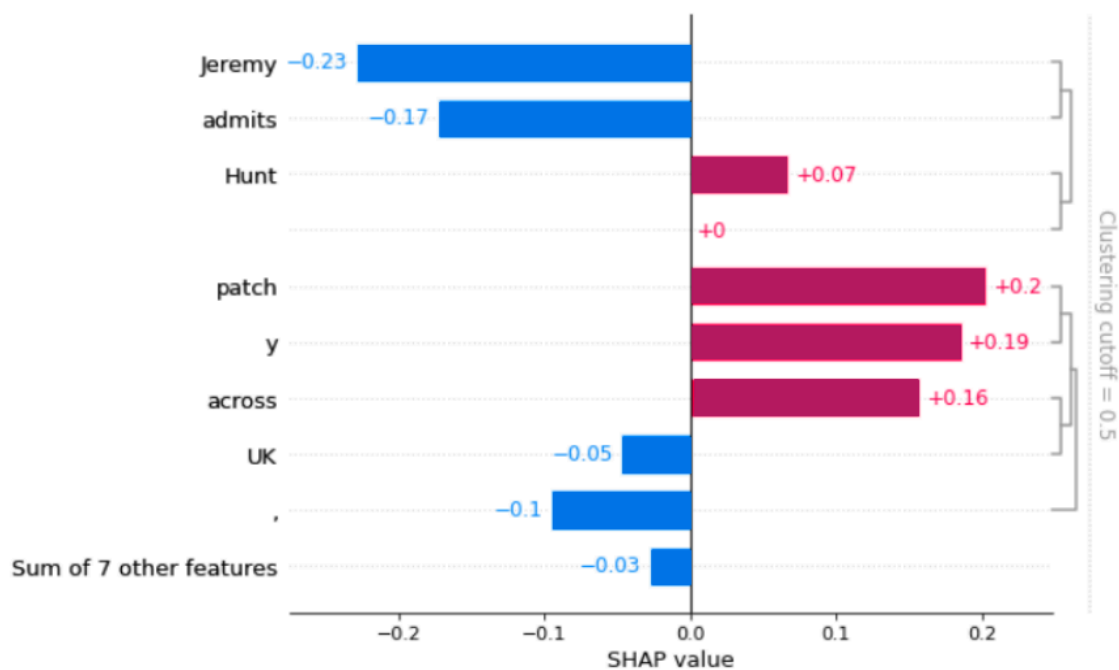
This visual representation provides information on the role each word played in the sentence's sentiment output. By looking at Figure 20, it was possible to identify the words that added a higher score of negativity to the headline such as 'patchy across', and those which added more positivity such as 'admits Jeremy'. I ran this example again but deleting this time the words 'admits Jeremy Hunt'. This new experiment resulted in a strong negative sentiment output that looked more coherent with the sentiment expressed by the news title. Entities like 'Jeremy Hunt' are one of several reasons why sentiment is misinterpreted by sentiment classifiers. Below, I list the cases where the tool and the gold labels disagree, categorising the mistakes according to their nature. This list provides useful examples for researchers to consider when using SA to interpret texts.

- **Entity**: As in 'Olympic legacy: school sports provision patchy across UK, admits Jeremy Hunt' (Press Association, 2012), entities carry sentiment. Sometimes, their sentiment is relevant to the analysis. However, this is not the case for this sentence as Jeremy is not the main subject of the headline but 'school sports provision'. Jeremy here added positivity, which had wrongly influenced the final sentiment output. To come to this conclusion, we replaced Jeremy Hunt with the names of two other British politicians mentioned in the corpus: Boris Johnson and Ken Livingstone. Despite the positivity of the word 'admits', both Johnson and Livingstone added significant negativity to the final score.

- **Syntax**: For this example 'The Essential Morning: the melancholy inheritance of the Olympics' (Globo, 2019), the sentiment output was positive, although this is clearly negative. The problem is that the first part of the sentence 'The Essential Morning' added positivity to the final score. This is the name of the newspaper's section where the news of the day is summarised in the mornings. Therefore, the first sentence is not part of the main topic covered by the news article. Two or more sentences separated by punctuation can cause some inconsistencies like this.

- **Semantic**: In 'Environmental legacy, the great debt of Rio Olympics' (Soares, 2016), although 'debt' was correctly assigned a negative label, 'great' was misunderstood. It was read as an adjective meaning 'good' instead of an adverb meaning 'big'. This mistake added high positivity to the sentence, influencing the final result.

- **Negation**: Negation is a huge challenge for NLP tools (Kassner & Schütze, 2019). In 'Olympic stadium will not be white elephant after London 2012' (Gibson, 2011) the word 'not' changed the meaning of the rest of the sentence to its opposite. Some models not only ignore this feature but also give to the word 'not' a high negative score, which leads to improper classification of the text.

- **Metaphors**: White elephants are among the most representative negative aspects of the Olympic legacies (Leopkey & Parent, 2012). The term usually refers to those infrastructures constructed for the games that were abandoned right after the event or remain in a precarious condition not serving the purposes for which they were initially planned. It, however, was not classified as negative by the algorithms in 'Letters: The true Olympic legacy is white elephants on our doorstep' (Primm, Wood & Rose, 2012).

- **Domain-specific words**: In the title 'Britain's Olympic legacy is a sedentary nation' (Conn, 2015), the word 'sedentary' carries a specific negative connotation. In the context of Olympism, a sedentary nation means that people are not playing sports as desired by the legacy plan. It means that the plan failed to inspire local citizens to be more active. In this example, however, the word sedentary was assigned with a positive label.

- **Sarcasm**: As even humans struggle to identify sarcasm, for machines the task is even harder (Farias & Rosso, 2017). In this news headline 'Martin Samuel: Tessa Jowell deserves an Olympic medal in utter madness' (Samuel, 2010), the expression

'deserves an Olympic medal' was classified as highly positive, ignoring the sarcasm in the sentence.

● **Objective statements**: For sentences like 'London 2012 Olympics will cost a total of £8.921bn, says minister' (Gibson, 2012b), it was difficult to assign a label as the sentiment of the word 'cost' varies depending on the context. A cost lower than expected is positive, for example. However, words like 'cost' and 'pay' were mostly classified as negative by the algorithms. 'Investment', however, was classified as positive.

A few words on objective statements is necessary. I have frequently been asked in conferences the following question: 'but aren't news supposed to be neutral?'. The principle of objectivity in journalism is associated with neutrality and impartiality (Ojala, 2021). In this context, it means that journalists should contemplate the different perspectives and sides of a story, avoiding positioning themselves in favour of one cause. As stated by Boudana (2016), impartiality is based on 'non-partisanship' and 'balance' between the divergent interpretations of events.

What sentiment analysis classifies, however, is sentiment discursively expressed in the text. The concept of neutral, in this sense, is not the same as the one used in journalism. For SA, neutrality is the presence of words that do not carry sentiment. As shown in the example above, however, determining what is neutral can be a challenging task, as even verbs like 'cost' can have an associated sentiment depending on the context in which they occur.

A last point to be mentioned regarding the use of sentiment analysis techniques is the effect of translating texts originally written in Portuguese into English.

3.7.5 Impacts of translation in SA

As discussed above, the use of SenticNet, SentiStrength, Vader and BERT to classify Portuguese text produced unsatisfactory results. Working with data in Portuguese consequently required automated translation into English. Although widely used and recommended in some cases for non-English SA (Araújo et al. 2016, 2020), machine translation impacts the scores and can lead to misreadings (Dashtipour et al., 2016). However, Balahur & Turchi (2012) evaluated the impact of machine translation from German, French and Spanish into English, with two being romance languages like Portuguese, and concluded that the noise existed, but the result was still reliable.

The chart in Figure 21 shows the comparison between labels produced using the four classifiers for the original news titles in Portuguese and the translated version in English. BERT trained on Amazon and Twitter datasets were compared individually with Portuguese Twitter BERT. The main purpose was to observe how often translation changes the sentiment of the sentences. The chart represents a total of 464 articles published in Portuguese divided into three sections. Blue represents the number of titles that kept the same label in both languages. In yellow are shown those titles whose sentiment changed in the opposite direction: from positive to negative or from negative to positive. On the top of the bars in red are those titles whose sentiment varied slightly from positive or negative towards neutral, for example.

Figure 21 – Portuguese (original) vs. English (translation). Variation of sentiment in the 464 news titles originally published in Portuguese by the Brazilian press and then translated into English
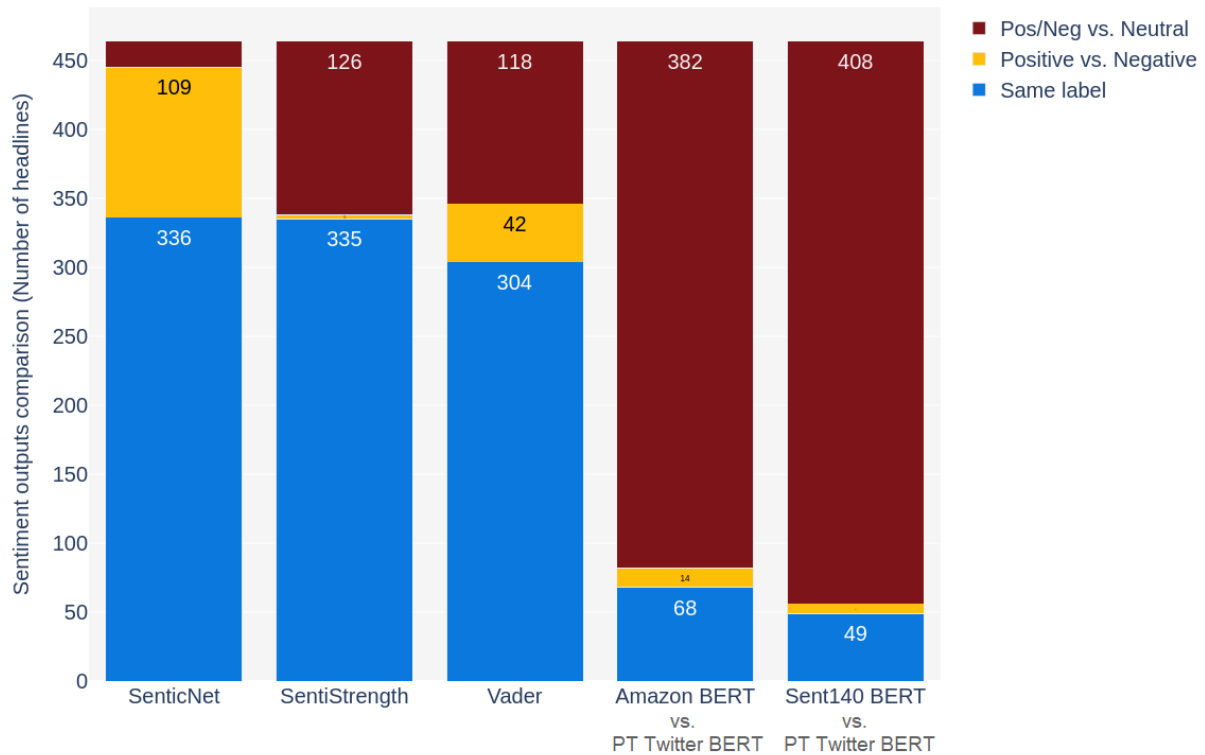
According to the chart, SenticNet and SentiStrength resulted in around 72% agreement between the labels assigned to the original and translated versions of the sentences. Vader presented slightly fewer matches, with 65% of the titles labelled the same. Here, I highlight the amount of titles whose sentiment varied from the opposite label, in yellow. SenticNet presented the highest disagreement at around 23%, followed by Vader with 9%. Amazon BERT and Sent140 BERT presented the highest variation towards neutral results. This occurred because most outputs from Portuguese Twitter BERT were 'neutral', as shown in Figure 21.

This experiment led to the conclusion that SentiStrength presented the best equivalence between Portuguese and English translated versions, presenting no change in 72%, slight variation in 27% and less than 1% of opposite results. The reason for this is that, as a lexicon based algorithm, it outputs a similar score for words in both languages, while the others are

affected by rules that take the syntax into consideration, producing different results. However, due to the lower efficacy of SentiStrength discussed above in the methodology, this classifier was not used. Nevertheless, this experiment provides a better understanding of the impact of the translation in Vader, one of the three classifiers selected for this study, and potentially other classifiers which rely on syntactic cues.

By looking closely at some of the sentences translated into English, some issues could be identified. In Table 7, there is an example of error in the first two headlines in the translation of the word *corte*. *Corte* in Portuguese can be both the verb 'to cut' or the noun 'court' in English. The first sentence should have translated the verb as 'cut' and the second as 'court'.

Table 7 – Comparison between the original sentences in Portuguese and their translations into English

Portuguese (original)	English (translation)
Brasil anuncia ao COI corte de R$ 900 mi no orçamento dos Jogos	Brazil announces the COI Court of R$ 900 mi in the games budget
Legado olímpico virou milhões em propina a 'amigos da corte' de Cabral, diz MPF	Olympic legacy turned millions in tipping 'Cut Friends' of Cabral, says MPF
Olimpíada é 'desculpa fantástica' para mudar o Rio, diz prefeito	Olympiad is 'fantastic excuse' to change the river, says mayor

The third sentence shows a mistake in the translation of the word 'Rio'. It should remain 'Rio' as the name of the city, Rio de Janeiro, and not be translated into 'river', the meaning of the word *rio* in Portuguese. These mistakes have impacted the sentiment of the news headlines and reveal what sort of issues can be faced by researchers when using this technique to explore large amounts of automated translated texts. This highlights the need for the development of more tools that work in languages other than English.

3.8 Discourse analysis: reading (beyond) the text

The methodology of this thesis combined both quantitative and qualitative methods. The use of these methods was not necessarily linear – quantitative first, qualitative second – but both have informed each other mutually in order to increase the analytical potential of this study. For the qualitative approach, this thesis resorted mainly to the principles and procedures of Discourse Analysis (DA), with a special focus on the French school. The reason for choosing this French tradition instead of Critical Discourse Analysis is due to the focus of the former on the aspects of ideology, deeply influenced by Lacanian psychoanalysis and Althusser's Marxism, which emphasise the social conditions for the occurrence of the discourse, and that are materialised in the text (Maingueneau & Angermuller, 2007).

Dominique Maingueneau (2004, p. 52), in his book *Analyser les Textes de Communication,* described eight essential characteristics of discourse. Discourse is:

(1) an organisation situated beyond the sentence: the discourse is not in the sentence – or the text – but it can be accessed through it.

(2) oriented: discourse is constituted in time and with a specific purpose (which can be deviated throughout the process).

(3) action: as postulated by Austin (1975) and Searle (1969), to say is to do. Therefore, discourses produce effects in the receptor.

(4) interactive: discourse is interactive not only in the sense that it involves the *I* and *You*, but also because it is dialogic. Every discourse responds to discourses previously produced.

(5) contextualised: it not only has meaning within a context, but also produces and modifies the context in which it was produced.

(6) assumed by a subject: every discourse has an *I* that is responsible for producing it – even when this is not obviously stated. Maingueneau (2004, p. 55) uses the example

of the sentence 'it is raining', in which the enunciator attests the veracity of the fact. However, this could have been framed in many different ways: *maybe it's raining*, creating uncertainty; *according to Paul, it's raining*, assigning the *I* to someone else.

(7) governed by norms: there is the need of legitimacy – within certain rules – for the discourse to be produced.

(8) part of an interdiscourse: a discourse coexists among other discourses.

What is relevant from these characteristics for this study is the understanding that discourses have objectives; are produced by a subject, with an intention, based on previous discourses, in certain historical and discursive conditions; and follow structures or 'norms' that guide how they can be formulated, depending on the circumstances in which they are communicated.

While this is not a thesis in linguistics, and therefore it does not deeply explore the different currents of DA nor focus on the theoretical foundations of this field, practical guidance on how to read texts and discourses has been used to inform the research practices conducted through close reading.

Consequently, the book *Discourse Analysis: Principles and Procedures* (Orlandi, 2015) has been an important source, with special focus on Chapter 3: Analytical apparatus. Orlandi drew her framework for analysis primarily from the studies of Pêcheux (1969), Foucault (1969) and Maingueneau (1984). The author described the main function of this analytical apparatus as:

> placing the said in relation to the unsaid, what the subject says in one place with what is said in another place, what is said in one way with what is said in another, trying to hear, in that what the subject says, what they do not say, but also constitutes the meanings of their words (2015, p. 59, author's translation).

Contrasting the said with the unsaid is a significant challenge for the analyst as it requires the ability to identify the 'silences'. Orlandi (2007) proposed a classification of types of silences or 'politics of silences' in the discourse: (1) The constitutive silence, which is intrinsic to any discourse, as it presupposes that in order to say something it is necessary to not say something else; (2) The local silence, which is censorship – when something is forbidden to be said. Comparing the news articles with government official websites and texts published by activists in blogs was useful for identifying this 'constitutive silence', as one exposes what is not being said in or by the other. Exercising the eyes to look for 'local silences' was fundamental due to the subtle ways in which it is manifested.

Another important point stated by Orlandi (2015) that has informed the analysis conducted in this study is that Discourse Analysis is inexhaustible. Therefore, it is not part of this approach's objectives to reach completion in relation to the empirical subject (Orlandi, 2015, p. 62). Also, deciding on what constitutes the corpus is part of the discursive properties (Orlandi, 2015, p. 63). This way, the researcher does not place themselves outside of history or ideology. Moreover, according to the author, the first process of de-superficialising the discursive subject consists of questioning: 'how was it said, by whom and in which circumstances?' (Orlandi, 2015, p.65).

Orlandi's analytical apparatus has served as a guiding principle for the close reading process. It helped to structure the reading by aiming at de-superficialising the discourse subject, accessing multiple discourses through the texts, looking for clues of silences (both constitutive as well as local) and situating myself as a researcher located in a historical time and constituted by ideology.

The way in which ideology is manifested through my actions as a scholar, as stated by Orlandi (2015), encompasses the choice of research topic, corpus, methods and analytical lenses. Therefore, being Brazilian, having studied journalism, having worked in the Rio Olympics, having studied activism in a Master's degree, having conducted this study in Digital Humanities at the University of London and many other contextual circumstances have influenced the conditions under which I approach the topic.

3.9 Online news: important considerations

When looking specifically at online news, there are some important characteristics of the medium that need to be considered, such as findability and the role of search engines, spreadability and circulation on social media, production logic and the commercial mechanisms of online marketing.

Findability and the role of search engines: users currently rely heavily on Google.com to access news content. As mentioned above in section 3.2 of this Chapter, according to the Ofcom News Consumption Survey 2020 (Ofcom, p. 56), in 2019, Google was the second most accessed website for news in the United Kingdom. It is possible to empirically attest that news articles also have an important relevance for Google's ranking, as they are constantly among the first outputs for a keyword search. Precise information on how this ranking system works is, however, not available.[51] What is important for this research is that, besides the widely studied dynamics of journalistic content production being framed by gatekeepers and agenda-settings, Google is a relevant player in the selection process of what deserves to be seen.

Spreadability and circulation on social media: the logic of circulation goes beyond the medium. Kümpel, Karnowski & Keyling (2015) provided a review of the research published on news sharing in social media, where some key aspects are covered: (1) social media has become indispensable for online news distribution and consumption and (2) online news outlets rely on reposting to increase their audience. Even though this thesis is not looking at social media data, this characteristic is fundamental for understanding how news content has been produced, which also has to do with the mechanisms of online marketing.

---

[51] After being accused of making money with third-party content produced by news outlets, Google defended itself by undermining the relevance of news for their business model: 'News queries on Search accounted for under 2% of total queries on Google Search globally in 2022'. Available on https://blog.google/supportingnews/#overview. Accessed on 14 August 2023.

Production logic and the commercial mechanisms of online marketing: platforms rely on clicks, demanding users to access their pages, to measure audience and produce profit (Scacco & Muddiman, 2016). An alternative model is to use paywalls and page blockers to limit access to the content to subscribers for a fee. Both models, either working together or independently, have an impact on the way news is produced and consumed. Relying on clicks can potentially tempt authors to adopt a clickbait logic. Alarmist headlines or those containing questions and generic affirmations tend to attract the attention of the audience more effectively (Scacco & Muddiman, 2016) and therefore increase traffic on the website. This has a first, perhaps more obvious impact on the production of news headlines, as they are the invitation to access the full content, but the text accompanying the headline can add further complexity, negation or explanation.

If we look at content shared on social media, due to the format in which it is usually presented on Facebook and Twitter, for example, users will primarily have access to the news headline, a picture and, sometimes, a brief introduction or subtitle. Accessing the full content to have the complete perspective on the text is a user's choice, which may be not an immediate one for many reasons, such as lack of good internet connection, lack of time or constraints on accessing news websites. Regarding this last point, it is important to reflect on how paywalls can limit access to the complexity, negation or explanation of the full text.

Looking at this logic of production, distribution and consumption helps us to understand that the phenomenon of a discursive construction of the Olympic legacy does not necessarily take into consideration news as a single, cohesive and complete product. It can be resignified, emphasised or ignored by readers as active distributors of the content via social media, or by Google's ranking systems. Or, thinking of the logic of production, it involves nuances produced by clickbait headlines that contribute to emphasise certain aspects of the stories to the detriment of others.

This thesis is, however, interested in the discourses produced in and by the texts, with both the conditions of production and consumption being relevant but not central dimensions of the analysis.

3.10 Conclusion

This chapter presented the pipeline design for building the research corpora, including data collection, processing and cleaning. Moreover, I reflected on the limitations of accessing and working with data in the Portuguese language. I also elaborated on the analytical framework used, including Discourse Analysis as well as a critical perspective on mapping social phenomena.

The scheme below shows how the different methods used in this research project worked together by moving from NLP at scale to close reading to respond to the research question: 'what is the legacy of the London 2012 and Rio 2016 Olympic Games as conceptualised through the British and Brazilian news media coverage of these events?'. In order to find aspects of Olympic legacy that received attention from the media, NLP, especially n-grams, was used. It was complemented by the extraction of named entities (NER) and the production of narrative network graphs, illustrating how these entities related to each other in the texts. A third stage of this methodology was the use of SA to help understand not only the topics covered but the ways in which they were framed. It further helped to reveal how emotionally weighted some topics were in comparison to others. That is the difference, for example, between saying that the legacy of Rio was 'negative' versus saying that it was 'apocalyptic'.

At the final stage, close reading was performed using the method of Discourse Analysis. The objective here was to add greater complexity to the analysis by identifying discourses, counter-discourses and inter-discourses that appear in the analysed texts. Returning to the example above, besides identifying the use of strong negative language, such as the word 'apocalyptic', it becomes possible to identify who used this term, in which circumstances and how this term was integrated within the text and the narratives constructed by each media outlet.

Figure 22 – Scheme showing how different methods work together to respond to the research question

Natural Language Processing (n-grams)

NER & Network visualisation of entities

Sentiment analysis

Discourse analysis
(close reading)
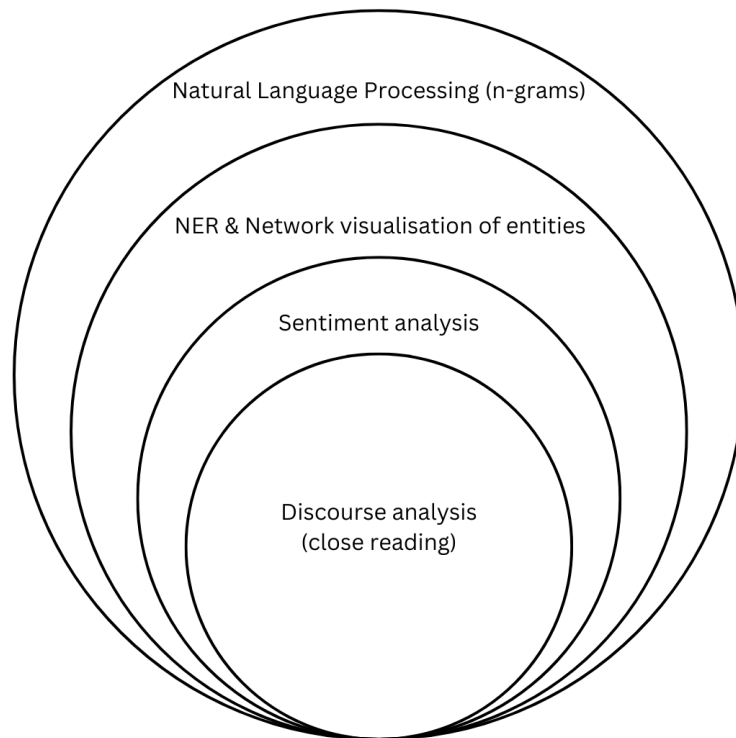
The image in Figure 22 shows four circles, one inside the other. The biggest one is NLP; inside it, there is a smaller one called NER and Network visualisation of entities, followed by SA and DA. It illustrates the approach to data analysis that characterises this PhD, moving from NLP at scale to close reading and discourse analysis.

Having outlined the methodology, in the next chapter, the results of the analysis performed on the news articles produced by the British media on the legacy of the 2012 London Olympics will be presented. The data will be discussed by drawing on the combination of both quantitative and qualitative methods.

Chapter 4: The British media narrative of London's legacy

Before diving into the analysis of the media coverage of London's legacy by the British media, it is important to look at two significant documents that set the tone for how legacy was constructed and narrated in both contexts: the booklets published by the London and Rio organising committees outlining their plans for delivering the Olympic legacy.

The London booklet was entitled 'Before, during and after: making the most of the London 2012 Games'. The document was published by the Department for Culture, Media and Sport (DCMS) in June 2008 – four years before the event. Rio's report was named 'Legacy plan' and was published by the Olympic Legacy Governance Authority (AGLO) in June 2017 – a year after the Brazilian edition of the Games. AGLO was created in 2017 (AGLO, 2017b, p. 13) to develop sustainable management of the Olympic venues. According to the document, it was intended that the body would mitigate the failures in the management of the infrastructure constructed for the event. The main objective was to make the venues usable, with ongoing activities and in an economically viable manner.

Looking at the titles, dates of publication and authorship, there is a clear difference between both cities in the way they understood the meaning of the word legacy in the context of Olympism. London's plan highlighted the broad temporal spectrum in which the event occurs – past, present and future – not only by including this perspective in the title but also by publishing the document prior to the Olympics. It is composed of five core promises: 'Making the UK a world-leading sporting nation; Transforming the heart of East London; Inspiring a new generation of young people; Making the Olympic Park a blueprint for sustainable living; Demonstrating the UK is a creative, inclusive and welcoming place to live in, visit and for business' (Department for Culture, Media and Sport, 2008, p. 1).

In the foreword, the Minister for the Olympics and London at the time, Tessa Jowell (2005-2010), emphasised that by watching previous Games, the lesson learnt was that legacy had to be planned early. Here, she extended the idea of legacy to the past by pointing out that

each Olympic Games event is a legacy of previous editions. Unlike the document published by Rio, London's report has a section entitled 'What we mean by "legacy"', in which legacy is described as a means for inspiring and motivating people and organisations 'to strive for their best' (Department for Culture, Media and Sport, 2008, p. 8). There is no clear definition of legacy in this section, though, as the title promised. The section on 'Legacy definition' in Rio's document is also generic; it consists of one paragraph stating that 'defining legacy is important' (AGLO, 2017a). Both reports show that there was a certain interest in defining 'legacy' although they do not in fact reach that objective, which means providing a clear definition of how they understand legacy and how this understanding would guide their legacy planning.

Throughout this chapter, I look at how the journalistic coverage of the Olympic Games reinforced or moved away from these official legacy plans. Moreover, by contrasting the media coverage with the government's official publications and also activists' blog posts, I aim to shed light not only on the topics covered but also on those 'silenced' by the press by identifying 'silences' in the media discourse, as conceptualised by Orlandi (2007).

4.1 Setting the agenda of the legacy

Daily news does not only inform the audience about events that have happened. Newspapers, television, radio and news websites decide on where the attention of the public should be focused by ranking articles according to their relevance. The role of organising events in terms of 'news values' is called agenda-setting (McCombs, 2004, p. 1). The main idea behind this concept is that readers can organise their own agendas by looking at these 'salience cues' (McCombs, 2004, p. 2), thus guiding public debate and therefore public opinion. In the seminal book *Public Opinion*, Lippmann (1922) stated that news media cooperate in the cognitive process of reading and interpreting the world. As established in section 2.3, framing legacy through the eyes of the media is an attempt to analyse its mediated construction as a

site of reference for citizens to form their opinion about public decisions that culminated in the hosting of this major sporting event.

As pointed out by Girginov, legacies become cultural capital as they are 'studied, documented, shared, and remembered' (p. 207). News platforms produce and reproduce the idea of Olympic legacy and they are understood in this chapter as active actors in the construction of this concept.

4.2 The British media narrative of London's legacy

Based on the list of bigrams presented in section 3.5, the following visualisation was produced to show the 10 most frequently occurring terms in articles published by the news media (British News Media Corpus), the government (British Government Corpus) and the activists' blogs (British Activist Corpus). The size of the internal circles represents how many times the term occurred.

Figure 23 – Most frequently occuring bigrams in news media, government and activist websites
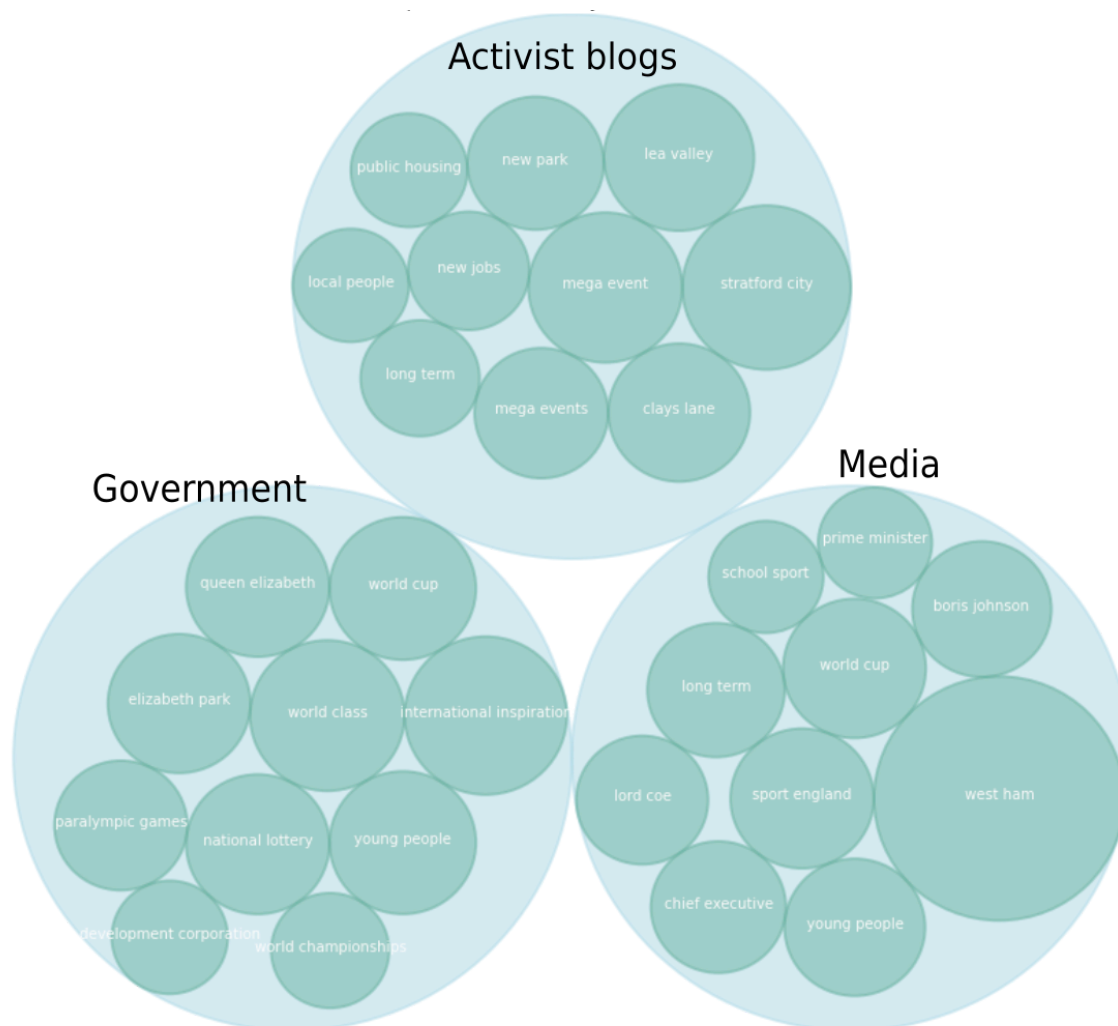
In the circle on the top, containing words mentioned in the British Activist Corpus, the most frequently occurring bigram is 'stratford city', followed by 'mega event', 'lea valley', 'clays lane', 'mega events', 'new park', 'new jobs', 'long term', 'public housing', 'local people', and 'local term'*;* for the British Government Corpus, the most frequently occurring bigram is 'international inspiration', followed by 'young people', 'world class', 'national lottery', 'world championship', 'development corporation', 'paralympic games', 'queen elizabeth', and 'elizabeth park'; for the British News Media Corpus, the most frequently occurring bigram, by far, is 'west ham', followed by 'world cup', 'sport england', 'boris johnson',

'young people', 'prime minister', 'school sport', 'long term', 'lord coe', and 'chief executive'. This list of trends generated from the three corpora was first used in a distant reading to give a sense of the most discussed topics and then combined later with a more qualitative approach of close reading to gain a deeper understanding of context.

The bigrams revealed a significant difference in the way the Olympic legacy of London was framed by the different media outlets. Among the most frequently occurring bigrams from news publishers are 'young people' and 'school sport', both referring to the promises included in the legacy plan of London. Promises number 1 and 3, entitled 'making the UK a world-leading sporting nation' and 'inspiring a generation of young people', included the engagement of young people in physical activities by increasing the offer of high-quality sports. The high co-occurrence of these terms (457 for 'young people' and 305 for 'school sport') positioned them as 4th and 9th in the list of top bigrams in the British News Media Corpus. Below, an image[52] generated from articles published by the BBC, using Voyant tools, shows the contexts in which the term 'young people' occurred.

---

[52] In this visualisation, you find the screenshot of the tool 'contexts' in Voyant. The central column shows the bigram 'young people'. The left-hand column shows what comes before the bigram in the text and, on the right, what comes after. This is helpful to understand in which context these words appear in the analysed articles. In the first line, for example, it is clear that the topic is 'barriers to young people' accessing sport activities.

Figure 24 – Visualisation showing the context in which the bigram 'young people' occurs

Left	Term	Right
for everybody. But there are still lots of barriers to	young people	in sport: money, gender and family - and how about just
funding regimes work for us to encourage more participation from	young people	coming through." Subscribe to the BBC Sport newsletter to get
Games will last for a long, long time especially for	young people	. "They will think 'gosh that was brilliant'. It'll be a
also heard from elite athletes how important it was for	young people	to access a broad range of sports so that they
said: "High quality school sport is vitally important. It provides	young people	with opportunities to improve their education, health and well-being
Twigg, Shadow education secretary "We need to be encouraging all	young people	to take part in sport; whether they enjoy the competition
with it many benefits, this emphasis can also deter many	young people	from taking part in sport at all. "We recommend that
competitive sporting opportunities for their pupils, to ensure that all	young people	feel able to take part in sport and enjoy the
grew up and home to the Olympic Stadium - to inspire	young people	to replicate her feats. Ohuruogu won silver in the 400m
see the benefits of sport and what it does for	young people	. "They understand very, very clearly the social and economic advanta…
Secretary Jeremy Hunt said the Olympics was inspiring millions of	young people	but insisted targets in education "don't always get you the
of team sport, and the values that can give to	young people	." The government would be bringing forward plans to make sure
government would be bringing forward plans to make sure more	young people	take part in competitive sport, Mr Hunt added. But shadow
the gap between schools and community sport. "We want more	young people	doing sport as a habit for life," said Culture Secretary
It will provide a whole range of sporting opportunities for	young people	through improved facilities and will I hope inspire them to
London's key Olympic legacy pledges - to inspire a generation of	young people	to take part in local volunteering, cultural, and physical activity
Olympic legacy, Lord Coe said the increase in numbers of	young people	, particularly with disabilities or impairments, involved in sport in Wales
have to look after the people who motivate children and	young people	." Dorset County Council chief executive David Jenkins pointed to lasting
It will provide a whole range of sporting opportunities for	young people	through improved facilities and will I hope inspire them to
years. "These investments will transform the local pitches where many	young people	have their first experience of sport," said Richard Lewis, who
Fund] will provide a whole range of sporting opportunities for	young people	through improved facilities and will, I hope, inspire them to
an east London firm to provide specially adapted bikes for	young people	with physical and learning disabilities, a pool lift at the
of school sport a week is a concern. "We want	young people	to do as much physical activity as possible in and

Notably, 'young people' appeared in several quotes from politicians such as Jeremy Hunt, the Secretary of State for DCMS (2010-2012), and Stephen Twigg, Shadow Education Secretary (2011-2013).[53] Over the years, there was a significant change in the way the term 'young people' was framed. After the Olympics, the drop in the number of 16 to 25-year-olds playing sport was one of the main topics highlighted by the media. The following piece, published by the Guardian in December 2013, sheds light on this issue:

> Despite the "inspire a generation" rhetoric used to justify the investment in the London 2012 Olympics, new official figures show that the number of 16 to 25-year-olds playing sport has gone down since the Games. (Gibson, The Guardian, 2013).

---

[53] The news articles also covered other legacy promises such as the projects to provide bikes adapted for young people with physical and learning disabilities.

Both 'young people' and 'school sport' were a response to the legacy plan published by the DCMS, and are investigated in more detail in section 4.6 and 4.7 of this chapter. However, the most mentioned bigram in the list of texts analysed did not receive as much attention in the London legacy official plan document as the aforementioned 'young people' and 'school sport': 'West Ham'. In the following sections, this bigram will be analysed separately combining quantitative methods with close reading.

4.3 West Ham Stadium

The destiny of the Olympic Stadium became one of the most controversial events in relation to the Olympic legacy of London. Initially, the disagreement on whether it should remain as an athletics venue or be handed over to West Ham United FC drew the attention of the media, with important voices like the Olympics Minister Tessa Jowell and ex-London mayor Ken Livingstone opposing the football club. The dispute between West Ham and Tottenham Hotspur FC over the Olympic Stadium and the threat of it becoming a 'white elephant' – a recurrent fear in recent Olympic history and one of the most representative negative legacies of the Olympic Games (Leopkey & Parent, 2012) – shed light on the site as a symbol of London's Olympic legacy, as highlighted in the headline 'Lord Coe: Olympic stadium will not be white elephant after London 2012' (Gibson, The Guardian, 2011).

By using the TIME Dashboard,[54] I traced how the discussion around West Ham evolved over time. The chart below shows word trends per year. The result presented is for the word 'ham' [55] mentioned by the BBC, The Guardian, Daily Mail and The Telegraph.

---

[54] Available at <http://cleopatra.ijs.si/dashboard/word_trends>. The website, however, presents recurrent failure as the server hosted by JSI is unstable.

[55] The TIME Dashboard does not support bigrams. 'Ham' was chosen as it is more specific to 'West Ham' than the word 'West' that appears in different contexts.

Figure 25 – Temporal distribution of the word 'ham' in the dataset (based on word count per year and per outlet)
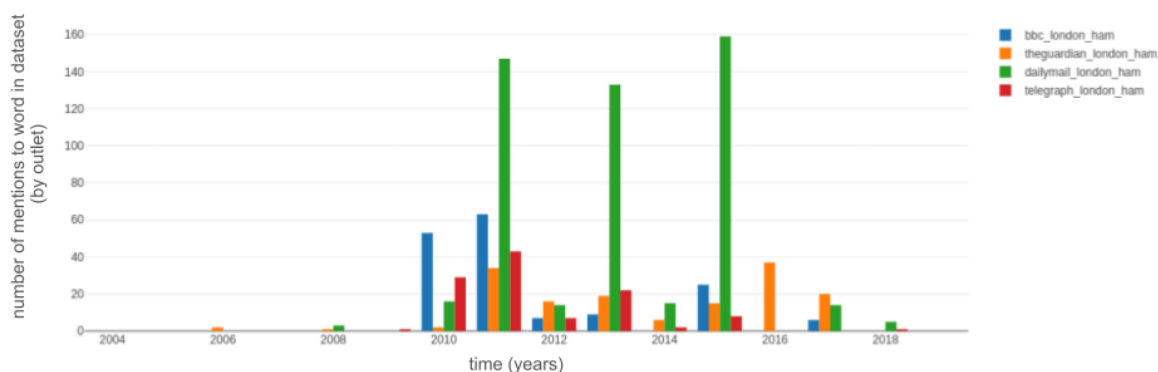
In the chart above, the Y axis shows the number of mentions of the word 'ham' in the dataset by outlet: BBC (blue), The Guardian (orange), Daily Mail (green), Telegraph (red). The X axis represents time in years. The chart shows that the word was mentioned considerably more by the Daily Mail than the other media outlets. In 2011, 2013 and 2015, in particular, mentions of 'ham' were notably high.

The list below shows the headlines of the articles represented in the chart presented above. It provides a similar analytical experience to the list produced with Voyant tools. The columns, from left to right, are: the name of the outlet (and the event covered by the article; the position of those articles in the results of the Google search; the title of the article; and the date of publication.

Figure 26 – List of news articles' headlines in which the word 'ham' is mentioned. Data includes date of publication, Google search rank and the news outlet

News Site	Search Rank	Words	Title	Date
london_bbc	182	ham	Tottenham 'would demolish Olympic Stadium after Games'	12/01/2011
london_bbc	276	ham	Tottenham 'would demolish Olympic Stadium after Games'	12/01/2011
london_bbc	265	ham	Athletics boss Diack slams Olympic stadium plans	20/01/2011
london_bbc	253	ham	David Bond: Original stadium plan not an option	21/01/2011
london_bbc	141	ham	West Ham beat Tottenham to win Olympic stadium backing	09/02/2011
london_bbc	237	ham	West Ham chosen as preferred Olympic Stadium tenant	11/02/2011
london_bbc	224	ham	Tessa Sanderson's London 2012 Olympic legacy fears	05/04/2011
london_bbc	147	ham	Adrian Warner	05/04/2011
london_bbc	272	ham	Adrian Warner: Sanderson breaks ranks to question legacy promises	05/04/2011
london_bbc	240	ham	Tessa Sanderson's London 2012 Olympic legacy fears	05/04/2011
london_bbc	282	ham	London 2012 Olympics Stadium row 'must not delay plans'	04/07/2011
london_bbc	271	ham	Sebastian Coe rejects 'white elephant' Olympic Stadium claims	19/10/2011
london_bbc	55	ham	London 2012: Olympic Games legacy promises under fire	08/11/2011
london_theguardian	211	ham	Olympic stadium: Don't betray the legacy of London's Games | Editorial	23/01/2011
london_theguardian	154	ham	What will be the legacy of the London 2012 Olympics? | Owen Gibson	31/03/2011
london_theguardian	58	ham	Olympic Park legacy official suspended over West Ham clash of interests	01/07/2011
london_theguardian	216	ham	Tottenham's fight with West Ham the tip of Olympic legacy problems | David Conn	06/09/2011
london_theguardian	53	ham	Lord Coe: Olympic stadium will not be white elephant after London 2012	19/10/2011
london_theguardian	60	ham	London's Olympic legacy must fulfil the promises made | Dave Hill	21/10/2011
london_theguardian	201	ham	Olympic legacy at stake in London's bid for 2017 World Championships | Owen Gibson	09/11/2011
london_dailymail	56	ham	Des Kelly: Spurs vandals musn't wreck Olympic legacy	15/01/2011
london_dailymail	112	ham	Tottenham and West Ham fight for the 2012 venue... will the Olympic stadium end up in foreign hands?	16/01/2011
london_dailymail	82	ham	Spurs fans protest over Olympic Stadium move but advisor voices 'white elephant' fears	16/01/2011
london_dailymail	84	ham	West Ham's plans for Olympic Stadium would turn it into graveyard says Harry Redknapp	21/01/2011
london_dailymail	141	ham	Tessa Sanderson warns Olympic Stadium chiefs: I'll take legal action unless you let me vote for West Ham or Spurs	21/01/2011
london_dailymail	50	ham	Charles Sale: 2012 Olympic Stadium legacy expert left two years ago	22/01/2011
london_dailymail	39	ham	Tottenham's hopes of securing London's Olympic Stadium in 2012 rise as research exposes legacy myth	05/02/2011
london_dailymail	51	ham	West Ham chief Karren Brady blasts Spurs: It would be a 'corporate crime' to knock down Olympic stadium	07/02/2011
london_dailymail	66	ham	Karren Brady is scaremongering, Tottenham chairman Daniel Levy hits back over Olympic Stadium 'crime' slur	07/02/2011
london_dailymail	145	ham	Spurs must win Olympic Stadium bid, says key London 2012 figure Mike Lee	08/02/2011
london_dailymail	120	ham	David Sullivan slams Spurs in Olympic Stadium battle	09/02/2011

The first documented use of the word 'ham' in the dataset is in just one article published by The Guardian in 2006, entitled 'No cut-price Olympic legacy for football's fat cats' (Conn, The Guardian, 2006). The article described the chances of a football club owning the Olympic Stadium after the event as 'extremely unlikely'. A year after London won the bid, Ken Livingstone, the Mayor of London, declared his disagreement with spending public money to construct facilities that would later be used by a 'rich' football club. Based on this understanding, it was proposed that the Olympic Stadium would remain as an athletics venue after the Games. This had not happened after the Commonwealth Games in Manchester 2002, as mentioned in this article:

> The London organisers are determined not to repeat what was done in Manchester where £120m of public money was spent building a permanent stadium for the 2002 Commonwealth Games - and then the track was removed so the stadium could be handed to Manchester City on terms regarded as startlingly generous. (Conn, The Guardian, 2006).

In 2008, the scenario was completely different. Two news articles appeared in the corpus in that year: 'London 2012 Olympic stadium may not have an athletics legacy after all' (Scott, The Guardian, 2008) and 'Spurs and West Ham bid for new homes as 2012 venues up for grabs' (Daily Mail, 2008). The main change in 2008, which would affect the destiny of the most discussed material legacy of the Olympics, was the election of the Conservative candidate Boris Johnson as Mayor of London. The effect of political changes throughout the long time between the bidding process until the execution of the Games is often responsible for significant transformations in the Olympic legacy planning. As stated by Girginov:

> The Olympics always takes place in a specific social, political and economic context that is not static but changes over the lifespan of a Games and after it has finished. Naturally, this context would exert different influence over the delivery of the event as previous editions of the Games have demonstrated. For example, the London Olympics bid was prepared and won by a Labour government, but delivered by a Coalition government which took a very different perspective on what the legacy should look like (2013, p. 199).

Unlike his predecessor, Johnson supported the idea that the private sector should manage the stadium after the Games, following the path of Beijing with the Bird's Nest, the Chinese Olympic Stadium for the 2008 Summer Olympics. According to an article published by The Daily Mail, one of the reasons for the shifting of purpose from keeping the space as an athletics venue to becoming a football stadium was the fear of the most prominent legacy of London 2012 becoming a white elephant. The article also reported the beginning of the dispute between two clubs to occupy the venue, West Ham United or Tottenham Hotspur, and the opposition of London 2012 chiefs, who defended the athletics facility as non-negotiable due to promises made to the IOC about the legacies of the Olympics.

In 2009, an article entitled 'London 2012: Serious questions remain about Olympic legacy' (Kelso, 2009, Telegraph) was published by The Telegraph. It highlighted the delay in decision-making about the future of the Olympic Stadium and attributed the 'delay' to the election of Boris Johson, whose government took time to decide on whether to create a new 'Olympic legacy company' that would replace the existing London Development Agency. According to the article, in February that year, Johnson gave up on changing the agency but the appointed Chair, Baroness Ford, would start working on the business plan for the Olympic Park only in October 2009, an eight-month delay.

From 2010, the number of articles mentioning the word 'ham' increased significantly. In February that year, the BBC published an article entitled 'West Ham's Olympic stadium plan angers Kelly Sotherton' (BBC, 2010b) echoing the opposition against the transformation of the stadium into a football venue. The heptathlon medallist, Kelly Sotherton, joined the ex-London Mayor and the Olympic Minister Tessa Jowell in defence of a national athletics venue.

In March, news articles discussing the preparation of West Ham United's bid to occupy the Olympic Stadium after the Games started to emerge in the British News Media Corpus. In the article 'West Ham and Newham Council eye bid for Olympic Stadium' (BBC, 2010), the club's co-owner David Gold used the expression 'white elephant' to raise concerns over the risk of having a stadium that would not be used to its full potential. To support his argument, he cited as an example the Crystal Palace Athletics Stadium, in south London, which he suggested was rarely used. Here, I highlight the use of the expression 'white elephant' by interested parties wanting to take over ownership of a symbol of London's legacy. It is used in the context of warning the government of the risks they would be taking by not considering West Ham's proposal. The use of such words emphasises the power that the expression carries, particularly with governments, which tend to fear accusations of delivering expensive unusable facilities after the event. It also reinforces Leopkey and Parent's (2012) statement that white elephants have particular importance in the definition of what a negative legacy is.

Dayan and Katz's (1992) definitions of the scripts for media events, as discussed above, were characterised by dramatic questions such as 'who will win?' (contest), 'will the hero succeed?' (conquest) and 'will the ritual succeed?' (coronation). The authors referred to the Olympics as a prime example of a contest script. However, when dealing with the Olympic legacy as a media event, the most significant question that shapes the media narratives, I argue, is 'will the event be prepared on time and efficiently?'[56] This question is equivalent to 'will the hero succeed?' in terms of building anticipation about whether the event organisers will manage to deliver the Games. By 'being prepared on time' I mean ensuring that facilities will be ready for use when the Games are expected to occur. 'Efficiently', however, entails more nuanced meanings such as the efficient use of resources; the construction of facilities that remain useful and used after the event; reduced environmental impact; positive social impact; among others. All these layers of efficiency corroborate the legitimating narrative that is idealised by the IOC and the governments involved (Poynter and MacRury, 2012, p. 315).

Still in March, the Guardian's article 'Uncertainty about Olympic legacy continues' (Hill, The Guardian, 2010) amplified the opinion of the Labour MP Kate Hoey, who argued that although she agreed with the private management of the stadium, there should be attention to the process to avoid excluding the local community from use of the space. She mentioned Wembley stadium as a place where, according to her, nothing happened during the day – an allusion to its 'white elephant' characteristics.

In June, Adrian Warner published the column 'Mayor aims for total control of 2012 legacy' (BBC, 2010). Warner argued that Boris Johnson's decision to concentrate responsibilities for the Olympic Park in his hands, by creating a new company run half by the Mayor and half by the government, had made the Mayor the main figure responsible for the consequences of the legacy. In Warner's view, handing the stadium over to West Ham seemed to be the easiest option to avoid accusations of producing, again, a 'white elephant'. The impetus to

---

[56] Dayan and Katz' students have already shown how their three scripts are not enough to explain every media event (Rothenbuhler, 1985).

concentrate power is commonplace in the administration of the Olympic legacy, as will be reinforced in the analysis of Rio in the next chapter. On the one hand, concentrated power provided political leaders with more control over decisions and therefore over possible failures. On the other hand, it is also a mechanism of self-promotion that contributes to their careers.

The articles covering the costs and possible uses of the stadium and the parties interested in bidding for it continued to be published during the year, always emphasising the decision to be made on whether track and field could share space with football. One actor in the dispute, UK Athletics, ultimately offered its support to West Ham bid in October 2010. This decision is explained by a BBC report (Warner, BBC, 2010) that Wes Ham had agreed to keep the running track. This decision also gave the Hammers the support of Sebastian Coe, who was initially against a football club taking over the stadium. Jumping to May 2022, ten years after the Olympics – and the year this chapter was written – this topic returned to the spotlight as the football club offered money to UK Athletics to give up on their right to use the venue (Rumsby, Telegraph, 2022), thereby allowing the Hammers to renovate and transform the space into a full-time football stadium.

4.4 Tessa Jowell deserves an Olympic medal in utter madness

In February 2010, the Daily Mail published the article 'Tessa Jowell deserves an Olympic medal in utter madness' (Daily Mail, 2010),[57] signed by the sports columnist Martin Samuel. The article provided a critique of the Minister for the Olympics, Tessa Jowell's, decision to support Lord Coe and UK Athletics' aim to make the Olympic Stadium an athletics venue after the 2012 Games. This article deserves particular attention as it seems to condense all the arguments presented in the different texts covering the destiny of the Stadium analysed in this chapter.

---

[57] The headline was later replaced by 'An Olympic medal in utter madness as Tessa Jowell aims to deliver Hammer blow to West Ham's future Stratford move'. The date of change is unknown.

As an article published in a column, the opinion of the author is openly expressed through the use of highly charged adjectives when referring to politicians. In the title, Martin Samuel made use of sarcasm to say that Tessa Jowell 'deserves an Olympic medal in utter madness'. While he used an adjective commonly attributed by men to women they disagree with, 'mad' (Packard, 2021), Ken Livingstone is referred to as 'pig-headed'.

The text starts by contextualising what a failed Olympic legacy looks like using Athens 2004 as an example. The author referred to the Greek Olympics as an 'extravagant three-week jamboree'. Samuel listed several facilities constructed for the Games, such as the velodrome, that were not being used anymore. A picture of London's Olympic Stadium during its construction is followed by a warning: 'Don't turn it into a white elephant: London's Olympic Stadium is still under construction, but what will happen to it after 2012?'

Figure 27 – Printscreen of the image and legend published in the article 'An Olympic medal in utter madness as Tessa Jowell aims to deliver Hammer blow to West Ham's future Stratford move' published by The Daily Mail on 23 February 2010

Don't turn it into a white elephant: London's Olympic Stadium is still under construction, but what will happen to it after 2012?

Looking closely at the image and its legend, it is possible to establish its function in the narrative as an illustration of the uncertainty. The grey colours, the machines working on a still deformed pitch and the cloudy sky shorten the boundaries between a scene of

construction or destruction. What the article does, and which can be found in the general news media narrative of the Games, is to take previous unsuccessful legacies as examples to project a possible future of failure. The image here contributes to developing this pessimistic, or dystopian, idea of what 2012 London's legacy could be. Such images are not produced in a vacuum. Those images are part of a narrative, which, as stated by Roland Barthes (1975, p. 238) and discussed in section 2.3, cannot be produced without referring to a system of units and rules. Or, as argued by Dominique Maingueneau (2004, p. 52), such discourses are interactive – and dialogic – which implies that they always refer to previous formulated discourses. Understanding these interpretative keys is fundamental for analysing an image of this kind as not merely an illustrative representation of the news article's content, but rather as a narrative device that exists in relation to previous discourses, and therefore is capable of creating ambiguity between this image of construction and destruction (or uncertainty).

Samuel continued this article with the statement: 'give the people what they want, not what you think they need'. The author disagreed with Jowell's proposal that a grand prix athletics stadium was needed. Samuel drew an argument opposing the will to 'reality'. For him, there is no doubt that football is much more profitable than athletics, despite the quality and performance of the athletes. Therefore, he argued, a Football stadium managed by the local council, Newham, and the local football club, West Ham, with an athletics track for sporadic events would be the ideal scenario.

To conclude, Samuel highlighted the desire of the IOC to be flexible and avoid reputational damage. This would mean reutilising existing venues, such as Wembley Arena, and controlling costs. The author, again, warned: 'there will be more White Elephants in our zoo than spectators to watch the athletics'.

This article illustrates the tone of the coverage, especially regarding the most frequently occurring bigram in the corpus: West Ham. It exemplifies what was observed using Sentiment Analysis (see Chapter 6), that is, the negative expectations of the British media in relation to the Olympics. Although the previous Olympics were held in Beijing 2008, the author used Athens 2004 as an example of what to avoid, as it is famously known as having

one of the worst legacies of the recent Games. This common exercise of looking at the past to predict possible futures shows the importance of legacy for the continuation of the Olympic Games. This means that in order to make 'legacy' work as a legitimating narrative, as stated by Poynter and MacRury (2012, p.315), it needs to have a history of effective delivery. Besides local concerns to make an appropriate use of public investments, the legacy of a city transcends its territory and transforms it into a case of 'dos and don'ts': what should and should not be done in the future, based on the experience of a particular host city.

4.5 Tottenham versus West Ham

In parallel to the athletics versus football story, a new relevant player is inserted into the discussion: Tottenham put in an official bid for the Stadium in September 2010. The dispute 'West Ham versus Spurs'[58] was referred to as a 'saga' by BBC Olympics correspondent Adrian Warner (BBC, 2010b),[59] in an allusion to its long duration.

In January 2011, the BBC published the article 'Tottenham would demolish Olympic Stadium after Games' (2011). For the football team, the seats were too far away from what would become the pitch and therefore not suitable for football. The text mentions UK Athletics chairman, Ed Warner, for whom Spurs' proposal to rebuild the stadium meant disregarding a tangible Olympic legacy. When using the expression 'tangible legacy' to defend the usage of the venue for athletics competitions, Warner shed light on the importance of materiality for the legacy discourse. The Olympic Stadium is not just one legacy of the Games but the most representative monument constructed for the event. In another article published by The Telegraph in the same month entitled 'Olympic Stadium will become "desolate graveyard if West Ham win battle with Tottenham", says Harry Redknapp' (Telegraph, 2011), the argument that a multi-use stadium was not wanted by West Ham was supported by its

---

[58] Spurs refers to 'Tottenham Hotspur Football Club'.

[59] Adrian Warner: Two Olympic sagas: Denials and broken promises
(cleopatra.ijs.si/dashboard/doc_display/london_bbc/108)

manager Redknapp, who said the proposal would 'kill the atmosphere at the stadium'. According to the article, for the president of the International Association of Athletics Federations, Lamine Diack, Britain's sporting reputation would be 'dead' if Tottenham's proposal to remove the track was accepted.

Looking at the metaphors used in the article helps qualify the story. In the headline, there is the construction of a morbid scene through the use of 'desolate graveyard'. Then, Redknapp said the proposal would 'kill' the venue's atmosphere and that keeping the track would be a 'nightmare' for West Ham. As mentioned above, for Lamine Diack, Britain's sports reputation would be 'dead' with West Ham's victory. The use of the words 'graveyard', 'dead', 'kill' and 'nightmare' shows the tone of concern expressed in the media and constitutes what I call the use of affective language to construct a dystopian imagery of legacy. By using the same interpretative keys applied to the image in the section above in the context of the dispute to inherit the Olympic stadium, the deployment of certain words must be read as a tool in a discursive dispute, in which the fear of delivering a negative legacy plays an important role. The political use of the word legacy, therefore, serves the interests of those wanting to promote uncertainty over decisions that do not privilege their gains.

4.6 'Young people' and 'school sport': a legacy for the next generations

The charts below show the frequency in which the words 'young' (Figure 28) and 'school' (Figure 29) occur over time in the British News Media Corpus.

Figure 28 – Temporal distribution of the word 'young' in the British News Media Corpus (based on word count per year and per outlet)
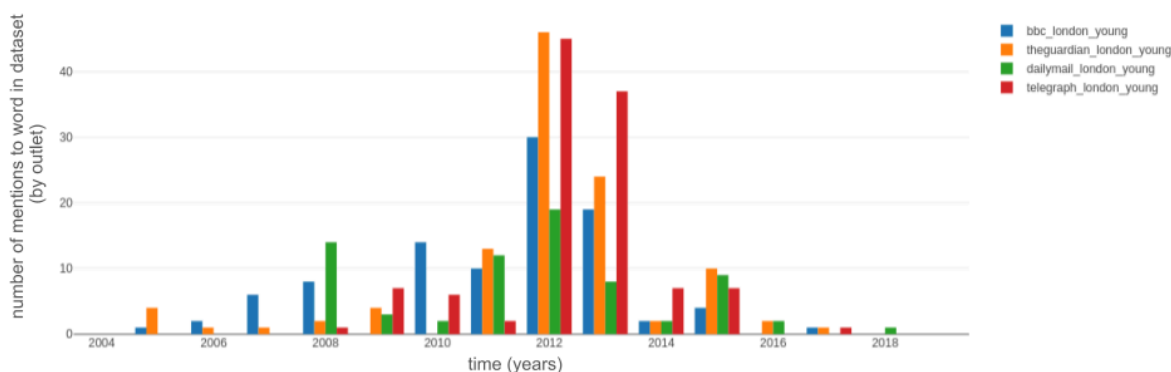
Figure 29 – Temporal distribution of the word 'school' in the British News Media Corpus (based on word count per year and per outlet)
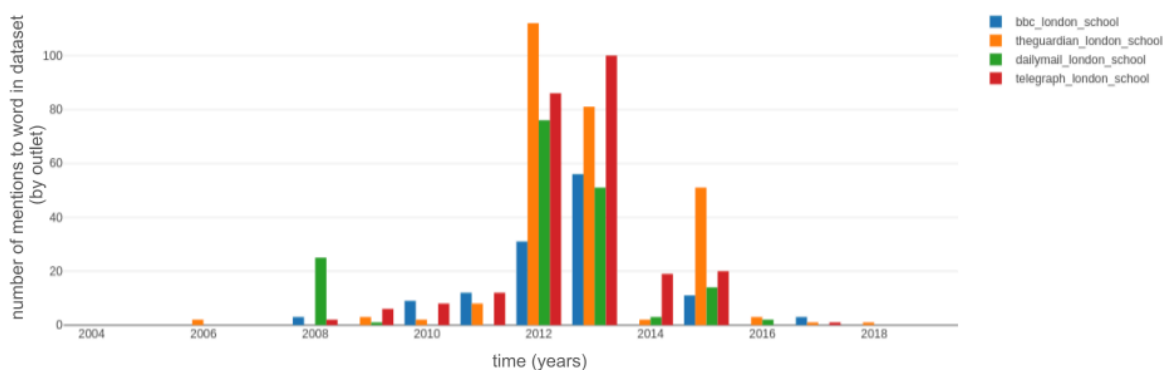
The Y axis shows the number of mentions of the words in the dataset by news outlets (BBC, in blue; The Guardian, in orange; Daily Mail, in green; and Telegraph, in red). The distribution in this visualisation is not normalised, therefore the imbalance in the temporal distribution of articles will affect the results. However, it is helpful to see the increase in the number of articles potentially covering 'young' topic in 2012 and 2013. The rise was considerably higher in those years for The Guardian and The Telegraph. It has, however, increased for all the analysed media channels, which shows the wide interest of the British media in discussing what legacy means for the young generations. The word 'school' follows

a similar pattern (see Figure 29). This mostly happens because 'school' and 'young people' are words within the same topic.

The news headline 'London Games would inspire athletes of tomorrow' (The Guardian, 2005), which belongs to the oldest article in the corpus to mention the word 'young', is useful to understand how the discourses concerning the impact of the Games on younger generations started. The article reports on the bid ceremony, where candidate cities presented their projects for hosting the event. That day, in the evening, London was declared the official host for the 2012 Olympiad. Tony Blair's speech is cited in the text, with its main highlight being the legacy for 'young people'. The main objective is described as inspiring young people to participate in sport and 'improving their lives as a result of the participation'.

Little further content on this subject was found in the corpus until 2008, the year of the Beijing Olympics. In August that year, a Telegraph article, 'Olympics: youth sports struggle as British win gold at Beijing Games' (Lusher, Telegraph, 2008), raised questions about the lack of investment in sports facilities for young people in the country, despite its gold medal success in Beijing. According to the article, elite athletes were receiving most of the funds while most people could not access a sporting space. Three days afterwards, however, the Daily Mail published the Mayor of London's response: 'Boris Johnson promises "tens of millions" to realise 2012 vision of a British Olympics legacy in 2012' (Crerar, Daily Mail, 2008), emphasising the government's commitment to invest in sports coaches for children. Johnson is quoted as saying that, from that moment until 2012, there would be a roller-coaster as British people, having watched and followed Beijing, were more aware of the issues and what he called the 'political disputes'. The Mayor not only clearly acknowledges the existence of political disputes for the legacy, but also introduces himself into the dispute when he touches on points that had been criticised by the media only three days before. By talking to the press about investing money in promoting sports for young people, he disputes the narratives with the media critique. His awareness of this narrative dispute is evidenced in the following paragraph:

> Mr Johnson said the success of Beijing had a 'big role' in persuading critics of the Olympics at home that they would be a good thing for the capital.

However, he accepted there would be 'ups and downs' in the run-up to 2012 as the Games became the hot political issue (Daily Mail, 2008).

The construction of Boris Johnson's image in relation to the topic of Olympic legacy will be discussed further in section 4.11, but it is worth taking a look at the imagetic representation constructed by the Daily Mail in this specific article, as can be seen in the image below:

Figure 30 – Printscreen of the image and legend published in the article 'Boris Johnson promises "tens of millions" to realise 2012 vision of a British Olympics legacy in 2012' published by The Daily Mail on 26 August 2008

He said London 2012 would see even more Team GB success if young Londoners were inspired to take up a sport and dedicate themselves to it.

Dedication: Boris Johnson. The London mayor said he wanted the 2012 Games to inspire a new generation of Londoners

Mr Johnson added: 'We are totally dedicated to it. I want to ring-fence London Development Agency money for sport and that means not just investing in facilities but also investing in the mentoring and the tuition that are so important for long-term success.

The Mayor of London is framed from bottom to top, which gives him the aura of a powerful man, a leader. The flags behind him suggest that this leadership goes beyond London, being

an international mission. In the legend, he is framed as 'dedicated' to his work. Although images are not the main object of analysis in this study they offer a contribution for contextualising the text and its embedded discourses. A series of immaterial legacies is invoked by the image, such as what Gratton and Preuss (2008) called 'interregional cooperation' and the idea of 'community spirit' that would be realised by inspiring a new generation. Moreover, such an image of a powerful political leader reinforces the concept of 'national self-confidence' (MacAloon, 2008, p. 2064), of a country that is a leader of and among other nations.

4.7 School sports: a response to the promises

Boris Johnsons's promises of investing in coaching and sports facilities for young people were heavily covered by the media, especially by The Guardian and The Telegraph. The article 'Government accused of "devastating" legacy with school spending cuts' (The Telegraph, 2010) refers to the promises made in the bidding process, notably that of inspiring a generation of young people, while reporting on the cut of £162 million of investment in physical education. The reduction of investment came to light in 2013, when the Guardian published the Active People survey, organised by Sport England, which revealed a decrease in the participation of young people in sporting activities as a result of the reduced funding (Gibson, The Guardian, 2013). The cuts reverberated over the years and in 2015 were cited by Tessa Jowell as the cause of a negative legacy. In 'Key London Olympics legacy "a failure", says Tessa Jowell' (BBC, 2015), the former Minister of the Olympics and, at that time, candidate for London Mayor, argued that although 'Inspire a generation of young people to take part in local volunteering, cultural, and physical activity' was one of the five promises, children were, to date, playing less sport than back in 2009.

In the following years, several articles were published raising this same issue, especially on the anniversaries of the London Olympics in the months of July and August. In August 2017, the BBC published the piece 'Olympic legacy: did £1bn after 2012 get any more people doing sport?' (BBC, 2017) highlighting that, despite a boost right after the Games, there was

no significant increase in sports participation. The same happened in August 2022, as The Telegraph published 'Were the London 2012 Olympics a failure?' (Rumsby, Telegraph, 2022b).

There seems to have been a misconception from the beginning about the 'inspiring a generation of young people' promise. While the media continued monitoring the delivery of the promised legacy, the erroneous formulation of the plan made its result unlikely to happen. A study published by Ken Green (2012) entitled 'London 2012 and sports participation: The myths of legacy' concluded, from a revision of government indicators, that neither the premise that young people were less engaged in sport activities before the Olympics nor the idea that hosting the Games inspires the audience to play sports are true. In another article published by BBC Future (Hammond, 2016), the author, Claudia Hammond, argued that 'rather than the Games failing to deliver the rise in participation, when you look at the evidence from past events, the failure was more the expectation it would happen at all'.

4.8 The government's contribution to the concept of legacy

To understand the media behaviour and decisions about how to frame the legacy of London 2012, it was necessary to compare media coverage with government official communications to understand the extent to which they influenced the media agenda. The following analysis was conducted on the British Government Corpus, formed from articles published by two government websites: uksport.gov.uk and queenelizabetholympicpark.co.uk. UK Sport defines itself on the website as the 'UK's trusted high-performance experts, powering our greatest athletes, teams, sports and events to achieve positive success'. Queen Elizabeth Olympic Park is the official page of the main space constructed for the Games. The total number of articles collected from these sites is 105, of which 68 were published by UK Sport and 37 by the Olympic Park. As mentioned in the methodology (Chapter 3), all the articles mention the words 'Olympic' and 'legacy'.

The most frequently occurring bigram encountered in the British Government Corpus was 'international inspiration'. These two words were mentioned together 119 times. The word 'international' alone is the 5th most frequently occurring (319 times, after 'sport', 'world', 'uk' and 'games'), while 'inspiration' was mentioned 123 times. To understand the context in which these words are mentioned, a close reading approach was conducted on the articles.

The bigram 'international inspiration' refers to the International Inspiration Programme, a project to promote sports in 'some of the most disadvantaged communities in the world'. According to the British Council,[60] the programme was delivered in partnership with UNICEF, UK Sport, Youth Sport Trust and the charity International Inspiration. The programme was a result of Sebastian Coe's promise in the 2005 bid to inspire a generation, which is in line with Promise 3: 'inspiring a generation of young people'. London was, in 2012, the first world city to launch an international legacy initiative. The programme's documentation stated that it aimed at targeting 12 million young people in 20 countries by 2012.

Many reports were published by UK Sport's website on the achievements of the programme, highlighting the countries affected by the initiative, such as India and Azerbaijan.[61] However, despite the government's choice to emphasise this programme over other legacy initiatives, what is not clear in the many articles published is what they mean by 'inspiring'. How can inspiration be measured? What are the objective indicators of successfully inspiring young people, or to be more (or less) precise, a generation?

---

[60] Information available at https://www.britishcouncil.org/society/sport/current-programmes/international-inspiration. Accessed on 11th May 2021. Archived on https://web.archive.org/web/20230606150030/https://www.britishcouncil.org/society/sport/current-programmes/international-inspiration.

[61] The programme reached 21 countries: Azerbaijan, Bangladesh, Brazil, Egypt, Ethiopia, Ghana, India, Indonesia, Jordan, Malaysia, Mozambique, Nigeria, Pakistan, Palau, South Africa, Tanzania, Trinidad and Tobago, Turkey, Uganda, UK and Zambia.

One of the articles, entitled 'Coe's London 2012 International Inspiration legacy promise reaches 25 million around the world' (UK Sport, 2014), is evidence of how abstract the communication of the programme objectives was. The article starts by highlighting that the programme 'has far exceeded [its] target to inspire 12 million children and young people across 20 countries'. It continues by explaining that the programme originated from Sebastian Coe's promise in 2005 to 'reach young people all around the world and connect them to the inspirational power of the Games so they are inspired to choose sport', an assumption that, as demonstrated above, cannot be sustained by evidence.

The text creates an expectation in the reader of finally understanding what 'inspiring' means when it presents an independent evaluation of the programme conducted by Ecorys UK. According to the article, the monitoring organisation's report concluded that the programme contributed to positive change in three 'levels of society': (1) 'over 25 million children and young people have been inspired'; (2) 'over 250,000 practitioners (teacher, coaches and leaders) have been trained, giving future generations around the world a stronger, more sustainable foundation to build local and national sporting systems and structures'; and (3) '55 national policies, strategies and legislative changes have been influenced'.

Again, the article argues that an independent organisation evaluated the programme and came to the conclusion that over 25 million people had been inspired. It does not, however, explain how. Among the benefits of being inspired, the report mentions 'healthier lifestyles, increased engagement in education, personal development and social inclusion'. It does not explain, again, whether the project invested in sports facilities, how many of them were constructed in different countries and how this could be related to inspiration.

Most answers to these questions are in the report published by Ecorys (2014). According to the document, the main focus of the programme was on the delivery of training on the management of public policies in the different countries. Through the technical guidance offered, the programme created connections between schools in the UK and abroad to

facilitate knowledge exchange; developed more of what they call 'safe spaces', mostly targeting sport activities for girls; and provided training to Physical Education professionals.

The programme failed in promoting more engagement in sports among 16 to 25-year-olds in the UK, as widely covered by the British media. But a second failure was in the communication of the programme. Although it was key to the legacy plan, the use of the word 'inspiration' brought an immaterial and abstract conception to the plan. While the government claimed the success of the programme by reaching millions of young people around the world, the media focused its attention on the lack of school sports facilities and young people involved in sporting activities in the UK. This evidence shows the British media focus on national rather than international legacy.

4.9 London's legacy entities: a political dispute

Returning to the news articles, network graphs were used to understand how key entities were represented in the media narrative. The graphs, as mentioned in Chapter 3, illustrate the connections between the entities mentioned together in the same article. The weight attributed to the edges is based on the number of times that two entities (nodes) were mentioned together in the same document. The more a specific entity occurs, the bigger it appears in the network.

By taking an overview of the network, it is possible to see the formation of three clusters, represented by three colours. One (in blue) revolves around Boris Johnson and Stratford, the second (green) surrounds the International Olympic Committee and the third (purple) is around the Paralympic Games and Great Britain. By looking at the centre of the network, two important figures can be easily recognised: Sebastian Coe, the London 2012 organising committee chairman, and Tessa Jowell, Minister for the Olympics (2005-2010).

Figure 31 – Visualisation of the centre of the network
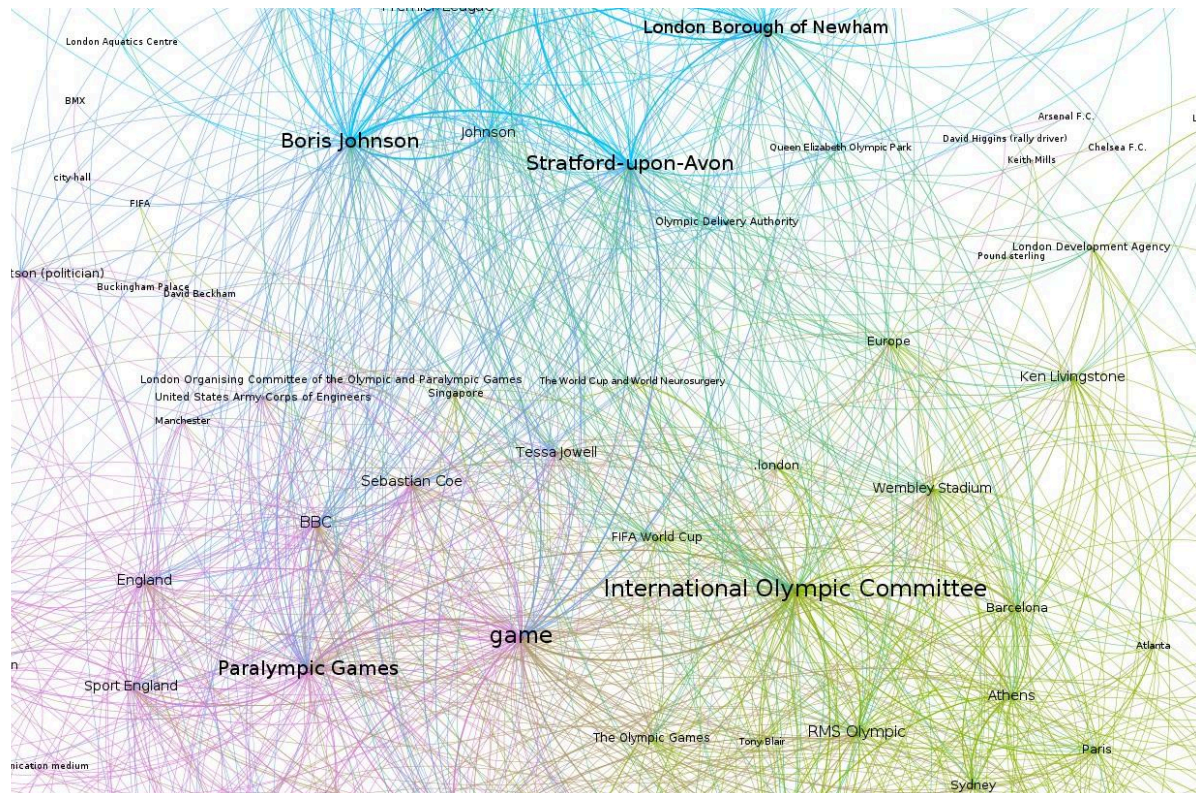
The centrality of both figures in the media's legacy discourse is clear after analysing some of the most frequently occurring bigrams in the coverage. Sebastian Coe presented London's bid on the 6th of July 2005 in Singapore. He was often referred to as the mastermind behind the International Inspiration Programme among other key legacies of London 2012. Tessa Jowell was the Minister for the Olympics during the Labour government. She was mainly mentioned as a defender of the use of the Olympics Stadium as an athletics venue, and from 2015 onwards, when she became a candidate for Mayor of London, as a big critic of London's legacy delivery. Jowell reversed her position from defender to attacker of the legacy during her career, maintaining her role as one of the most relevant political figures for London's legacy, even though she was not involved in the delivery after 2012. From 2010, when the Conservative David Cameron was elected Prime Minister, the position of Minister of the Olympics was held by Jeremy Hunt, who appears much less connected to the other entities and, therefore, at the periphery of the network.

By analysing Hunt's presence in the corpus, it is clear that his name was much less associated with the legacy narratives than his predecessor. Moreover, Hunt was likely less present because Boris Johnson, as Mayor of London, took on the role of manager of London's legacy. Jeremy Hunt, when mentioned by the media, appeared as a passive figure, assuming some of the negative consequences of the legacy plan, as in 'Olympic legacy: school sport provision patchy across UK, admits Jeremy Hunt' (The Guardian, 2012). Most of his mentions refer to his attempt – and failure – to increase sports participation across the country, especially among young people.

Despite the media coverage of his speeches promising the inclusion of more hours of sport in children's and teenagers' lives, in 2015 he was mentioned in the article 'Olympic legacy failure: inspiring London 2012 message has become a millstone' (Gibson, The Guardian, 2015) as unsuccessful in delivering it: 'At a valedictory press conference in Westminster the day after the Games, the then culture secretary, Jeremy Hunt, became visibly annoyed at a suggestion from reporters that his government would fail' (Gibson, The Guardian, 2015).

By tracking the last articles mentioning Hunt in the corpus, it is clear that he became strongly associated with the failure of this promise of inspiring a new generation. In the context of legacy, Jeremy Hunt assumed a negative connotation, which will be further explored and analysed in more detail with the use of SA techniques in Chapter 6.

4.10 Past and future examples: the cyclical legacies in the media

An article published by the BBC in 2004, months before London was officially elected the host city of the 2012 Games, had already initiated what would be the direction of the media coverage in the following years. Entitled 'UK urged to prepare for Olympics' (BBC, 2004), the article raised concerns about the delivery of an efficient legacy. To construct its argument, it made use of past examples of failure. It mentioned Atlanta 1996 as a place where house prices soared while new promised houses were not built. It also cited Sydney 2000 as a city struggling to make its Olympics facilities usable.

Another article, 'Olympics 2012: what will happen when the crowds leave?' (Telegraph, 2007) referred to Barcelona and Sydney as negative examples of under-used former Olympic villages, something that London should avoid. The world is full of past negative examples of legacy – having Athens 2004 being called the 'white elephants Olympic city' by Papanikolaou (2013) – and the media has tended to focus on them while covering the preparation of the future event. It is possible to visualise how this phenomenon occurs by looking at the green part of the network, in the image below.

Figure 32 – Visualisation of the cluster composed of geographic locations
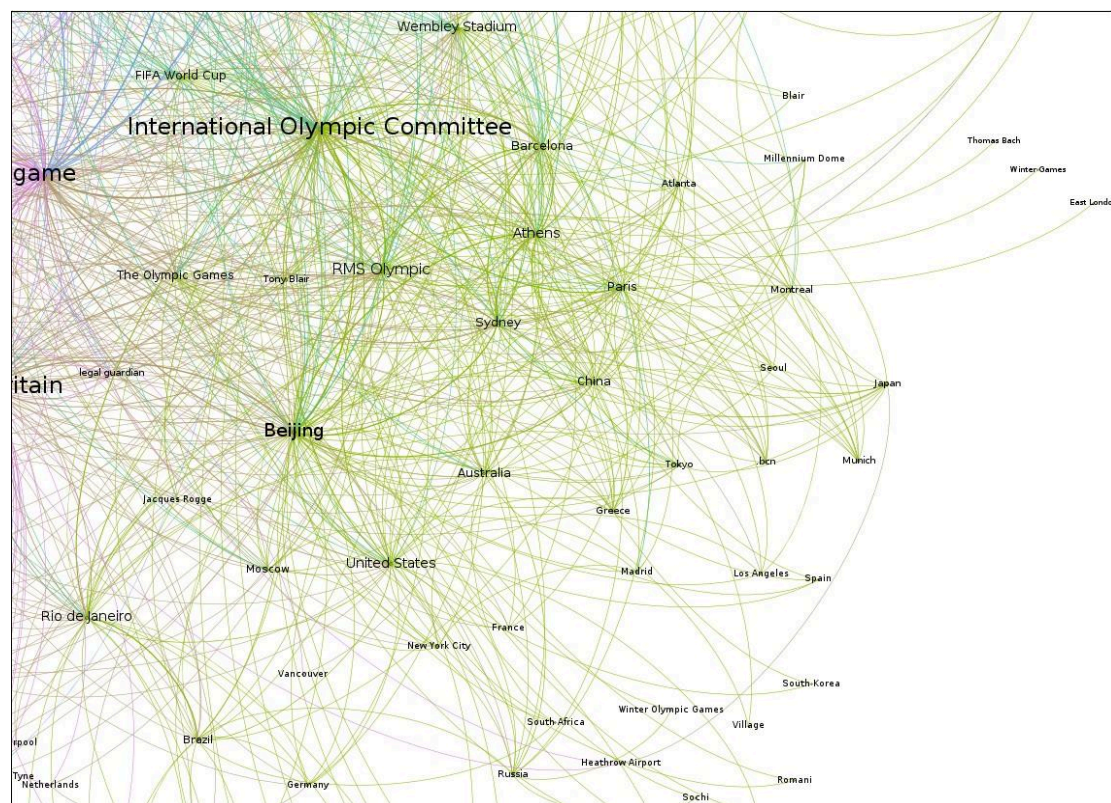
The connection between many different past and future host cities shows how often they are mentioned together in the news articles covering London's legacy. Beijing is a central entity, whose under-used stadium was referred to as 'essentially a tourist attraction' (Bond, BBC, 2013). Besides Beijing, the network includes the entities United States, Sydney, France, Russia, South Africa, Seoul, Montreal, Paris, Atlanta, South Korea, Sochi, New York City,

Brazil, Germany, Moscow, Rio de Janeiro, Spain, Munich, Japan, Barcelona, Netherlands, Vancouver, Madrid, Greece, Tokyo and Los Angeles.

However, the media not only looked backwards, but once the Games finished, future cities were mentioned in comparison to London's experience. In July 2016, The Guardian published 'The trouble with the Olympic Games' (Engel, The Guardian, 2016). The news outlet took the period of the Rio 2016 Games, and therefore the 4th anniversary of London 2012, as an opportunity to evaluate its legacy. The author, Matthew Engel, considered himself an 'Olymposceptic'. He argued that one of the main issues with the event was the fact that it was planned by one prime minister, Tony Blair, and delivered by another one, David Cameron. About the future, he argued that:

> Brazil has been in ferment. Tokyo, the 2020 host, is restless. Boston and Hamburg had to stop bidding for 2024 because of protests. By 2030 it will probably be only egomaniac leaders, mostly unelected, who will even dare to get involved. (Engel, The Guardian, 2016).

The way the arguments are concatenated in this article illustrates how past and future are often intertwined in the media discourse. Sometimes, the objective is warning the audience of possible negative results in the future; at other times, it is reflecting on their country's own experience before the next Olympics arrives.

As discussed previously in the context of white elephants, cities are constituted by dreams, but also by fear (Pinder, 2002). Having the city as the terrain in which legacies come about, those dimensions of desire/uncertainty are intertwined with the concept of legacy. It is common in the legacy discourse to make use of past negative examples to build threats, as in the case of the Olympic Stadium, or to simply keep the warning on the public agenda, positioning the media as a supervisory body of public investment. Those threats can also serve political careers, as we shall see in the next section.

4.11 Boris Johnson: a legacy career

> Zoom out a bit and the most tangible legacy of 2012 is the career acceleration of Boris Johnson, the kind of politician who would steal Judi Dench's Olympic skiffle board and outsource it to a Bosnian crime ring if it suited his personal advancement. (Ronay, The Guardian, 2022)

The chart below (Figure 33) shows the number of mentions of the word 'Boris' in news articles over the years. Johnson remains relevant from 2008 until 2020, with spikes in 2012, 2013, and 2014.

Figure 33 – Temporal distribution of the word 'Boris' in the dataset (based on word count per year and per outlet)
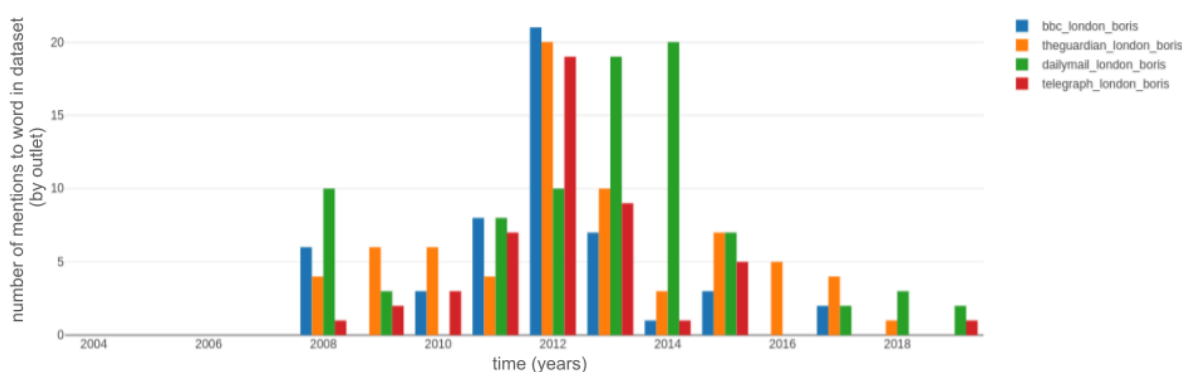
Since 2008, when Boris Johnson became Mayor of London, news articles started to be published about his promises for the Olympics, as in 'Boris Johnson promises "tens of millions" to realise 2012 vision of a British Olympic legacy in 2012' (Crerar, Daily Mail, 2008). From 2009, the press began to highlight Johnson's participation in the decisions regarding the Olympic Stadium, with 2011 being the year with the majority of articles covering the topic. In 2012, the delivery of facilities and programmes brought the name of Boris Johnson closer to positive outcomes, or at least attempts at positive outcomes, as in

'Boris Johnson urges two hours of PE a day*'* (BBC, 2012) or 'Boris Johnson announces cycling legacy plans*'* (Brain, BBC, 2012). The Guardian also reported that 'Boris Johnson takes charge of Olympic Park's future' (Gibson, Guardian, 2012).

The year after the Olympics, Martin Samuel's column celebrated the successful conclusion of the dispute around the Olympic Stadium in 'Martin Samuel: West Ham saved the Olympic Stadium from the idiots who built it, no wonder Karren Brady is smiling' (Samuel, Daily Mail, 2013). In the Telegraph, there was a note of celebration: 'Boris Johnson hails 'momentous' West Ham Olympic Stadium deal*'* (Telegraph, 2013). In 2014, the positive outcomes connected to Johnson's name kept being reported in the media, as in 'Prince Harry and Boris Johnson enjoy a playdate at the Olympic Park' (London, Daily Mail, 2014). In the following years, the occurrence of the name Boris Johnson reduced significantly in the British News Media Corpus.

Barney Ronay's article for the Guardian, published in June 2022, refers to Boris Johnson as a politician who had a boost in his career as a consequence of the Games. The bigram 'Boris Jonhson' was mentioned 447 times in the news cluster, being the 5th most frequently occurring in the British News Media Corpus. Johnson, who was the Mayor of London during the Olympics, was mentioned in the media on many different occasions. By looking at the network, though, it seems clear that a big part of his success in the corpus was due to his intervention in the destiny of the Olympic Stadium – based on the strong connection between 'Boris Johnson' and 'West Ham United F. C'. As discussed above, Johnson was an important figure in the decision to hand the stadium over to West Ham. He positioned himself in the middle of the dispute, becoming a key actor in the story.

Figure 34 – Visualisation of the cluster showing the relationship of the entity Boris Johnson with West Ham United F.C.

The way Boris Johnson appropriated the legacy discourse in his favour can be identified in some of the articles published by the media. In 'Will getting Brexit done create another baby boom in the UK?' (O'Carroll, The Guardian, 2019), Johnson used the success of the London Olympics, represented by the 'euphoria' of the Games, to defend a 'positive' result of Brexit. In his words 'romance will bloom across the whole nation' after the country leaves the European Union. Johnson makes reference to a 'baby boom' that, according to him, happened after 2012 as a result of the Olympics, something that had not occurred in the UK since 1967, when the country won the FIFA World Cup. However, as the same article clarifies, Boris Johnson's argument is false as there was no baby boom in 2013. In fact, the number of babies born that year was slightly smaller in comparison to 2012.

Although false, Johnson's argument spoke to people's memories of joy and celebration during the Olympics. It made reference, although sustained by wrong data, to a feeling that

people in London were aware of. This moment of happiness is compared to the effect of a drug by Moore (2013). She argued that during the Olympics all the wider political issues were suppressed by the magic of the Games. And this nostalgia was used by Boris Johnson to communicate a state of feeling in a way that people could relate to and understand.

In December 2019, now as the UK Prime Minister, Johnson tweeted his desire to host the football World Cup in 2030. 'I want it to show our national confidence as we get Brexit done', wrote Johnson. This strengthens the argument that Boris Johnson used sporting events as discursive tools to support his political advancement by promoting an emotional connection with his readers.

Figure 35 – Screenshot of Boris Johnson's tweet on hosting the World Cup in the UK
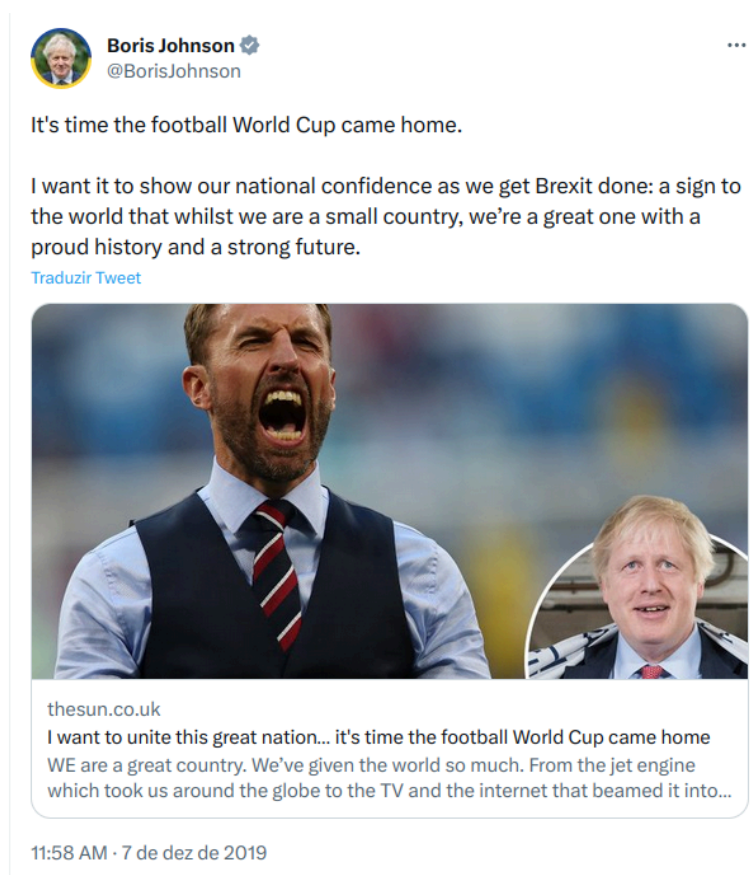
Behind this simple tweet, Johnson operated several intra-discourses, which Dominique Maingueneau (2004, p.52) called the interactive dimension of the discourse. First, Johnson alluded to the country's pride in being the inventor of football when referring to bringing it 'home'. Then he mentioned 'national confidence' as a necessity in post-Brexit times, followed by the representation of the UK as a 'small country'. In this sense, he counters negative discourses of the UK having become small in the geopolitical landscape after leaving the European Union. Finally, he connected the past – a nostalgic 'proud history' – to the desirable 'strong future'. By doing this, Johnson projected a feeling of belonging and success promoted by the recent London Olympics – which he claimed to be mainly responsible for delivering – onto a future of success in times of uncertainty and fear. This reveals how mega sporting events contribute to the political discourse and, more specifically, to politicians' careers, which is also evidenced in the Brazilian case to be analysed in Chapter 5.

4.12 Media silences: the activist blogs warnings

With the objective of analysing not only what is explicitly narrated in the media discourse but also those topics that did not receive much attention, activist blogs that raised concerns about the legacy of 2012 Olympics were incorporated into the British Activist Corpus. As an author, I was particularly motivated by previous connections with social movements focused on urbanism, as mentioned in the introduction of this thesis, as well as by the previously mentioned understanding that the city is 'an object of utopian desire, as a distinctive place of belonging', as postulated by Harvey (2012, p. 16). In this sense, the city is the territory in which legacy is manifested and semantically disputed by the different stakeholders, having the voices of local residents often being silenced. But as an object of desire, it is also a site of resistance, where contrasting ideas of what development means have to inhabit the same space.

Articles published by gamesmonitor.org.uk were collected using the UK Web Archive, as the blog was not available online on the live web at the date of data collection.[62] Data from citymonitor.ai was collected using the same method applied to most media outlets, via Google search. Games Monitor, in this corpus, is composed of 80 articles and City Monitor of 29.

Among the top 15 bigrams retrieved from the activist blogs are 'stratford city', 'mega event', 'lea valley', 'clays lane', 'new park', 'new jobs', 'long term', 'public housing' and 'local people'. As the first step for analysis, this list was compared with the list of top bigrams for news and government official websites. 'Long term' is the only bigram found in the British Activist Corpus that also appears among the 15 most frequently occurring in the British News Media Corpus. Others like 'local people' and 'lea valley' are mentioned, but only when considering the top 50 bigrams. An equivalent to 'public housing' that appears in the media cluster is 'new homes'. In comparison with the government cluster, there is no overlap among the top 40 bigrams. 'Long term' and 'local people', however, occur among the top 50 bigrams.

What this list of bigrams reveals is the local characteristics of the activist discourse. Their concerns targeted the impact of the regeneration programme for East London, mainly around the Olympic Park in Stratford. This was a response to two of the promises of London's legacy plan: Promise 2 – Transforming the heart of East London; and Promise 4 – Making the Olympic Park a blueprint for sustainable living.

In this context, Clays Lane became a symbol of the impact of the Olympics on the local community. The housing estate, built to offer assistance for people in a situation of vulnerability, was demolished for the construction of the Athletes Village and the approximately 400 residents had to move out. This is only one of the consequences of bringing the Olympic Park to East London.

---

[62] In July 2022, the website was revisited and it now offers an 'archived' version in HTML on gamesmonitor.org.uk/archive.

According to Poynter (2013) the transfer of the Olympic Park to the private management of Triathlon Homes and a consortium formed by Qatari and Delancey (two international investment companies) significantly affected the housing market in Stratford:

> Dependence upon private capital to deliver public gain has produced socially regressive rather than progressive outcomes for local communities. Public investment has generated value creation that is concentrated in pockets of gentrification: public agencies engaged to deliver legacy have been unable to secure its socially equitable distribution. (Poynter, 2013, p. 124).

Many properties in the area adopted a hotel style, targeting international business travellers. The impact of these decisions on the lives of East London residents received much less attention in the media when it comes to the articles published about the Olympic legacy.[63]

The topic of regeneration received more attention from The Guardian than the other outlets as illustrated in the image below. The image shows a spike in mentions of the word 'regeneration' in 2012, by The Guardian (in red).

---

[63] However, it is not possible to affirm that these topics were not covered in different contexts. For example, the article 'Going for GOLD! How east London homes have shot up in value since the 2012 Olympics with the price of properties in Waltham Forest more than DOUBLING' (Wright, Daily Mail, 2021) reports on how the 2012 Olympics' 'regeneration' of East London had directly affected housing prices, with increases between 60 and 100%. The article, however, does not mention the word legacy.

Figure 36 – Temporal distribution of the word 'regeneration' in the dataset (based on word count per year and per outlet)
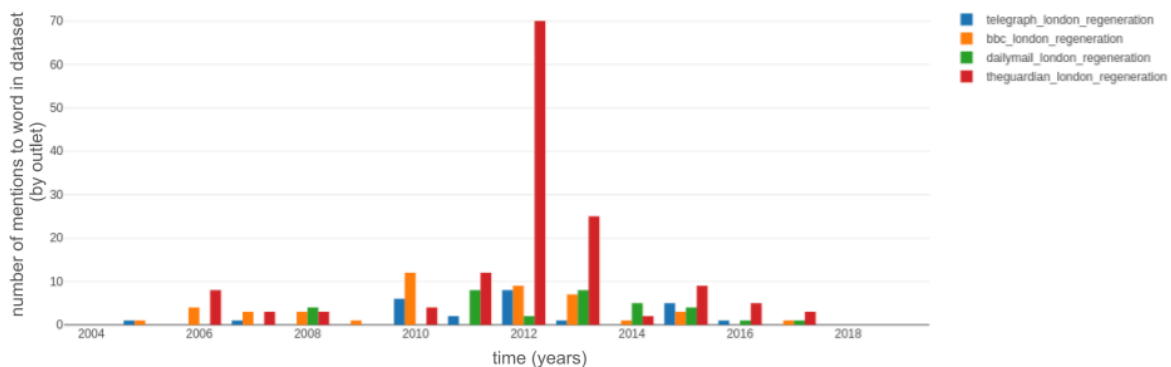
The fears of East-London residents, as in 'London 2012 legacy, the Olympics has cost me my home' (Grandjean, The Guardian, 2012), shared space in The Guardian with less critical coverage of the 'regeneration' plan, as in 'The Olympic legacy: creating a new community for London in Stratford' (Meredith, The Guardian, 2012). The latter, although published in the section 'News', was written by Dale Meredith, executive director of Triathlon Homes. There is no information on the page about whether the company paid to have this piece published. However, the language used in the article undoubtedly has a much more commercial than journalistic purpose: 'With a good number of family-sized homes, the East Village will provide much-needed affordable housing for east Londoners and, for many, this will be the true Olympic legacy', wrote Meredith.

Another example of the positive view of The Guardian regarding the redevelopment plan of East London is 'Stratford's Olympic legacy: It's been far better than I thought' (Hill, The Guardian, 2012). The article highlights the positive atmosphere provided by the Games for the residents of Stratford, despite their scepticism before the event. It also mentions the economic boost as a result of the Olympics, with some shops seeing their profits go up by 50%. A resident is quoted as saying that, although she used to think that the area was not good for living, it was getting better after the event, with 'buildings going up everywhere'.

The article contrasts the positive view at the end with the opinion of 'a group of local artists' who affirmed not having seen inclusion or jobs for locals as a result of the redevelopment.

One hypothesis that can be formulated based on the exploratory quantitative analysis in combination with close reading is that the effect on housing costs is not considered under the broader understanding of what legacy means. Looking specifically at the Daily Mail's coverage of the legacy, most of the topics that gained attention in their publications were related to the destiny of the Olympic Stadium and the West Ham versus Tottenham conflict. This way, it is possible to argue that the Daily Mail's concept of legacy refers to the material, tangible and polemic structures constructed for the Games. The Guardian, on the other hand, touched on the regeneration of East London from the perspective of urban development, giving space to the voices of business to the detriment of the local residents' struggles. The particularities of each news outlet will be covered in more detail in the following sections.

4.13 International inspiration versus local aspiration

Nauright (2004) described how sporting events have been used to reinforce local and national identities as a result of globalisation. The author also highlighted how international development strategies have become a key component of sports mega events (Nauright, 2004, p. 1325). By comparing the narratives of the press with the government discourse and the activists blogs, this section will explore the different ways the local, national and international dimensions of legacy were addressed.

The media coverage of London's legacy contrasts with the much more abstract and broader bigram found most commonly in the texts published by the British government: 'international inspiration'. Articles published by uksport.gov.uk were previously revealed as focused mainly on The International Inspiration Programme. While the media was more focused on internal, national issues, such as the lack of sports activities for young people or the economic impact of the future of the Olympics Stadium, the government was targeting international audiences. The choice of the word 'inspiration' further references a much more immaterial

and abstract idea of legacy that contrasts with the very concrete discussion around the Olympic Stadium.

Looking at the bigrams obtained from the British Activist Corpus, the concerns are shown to have been more local, primarily targeting challenges faced by citizens of East London. Among the main bigrams are 'Stratford City', 'new jobs' and 'public housing'. This community-focused approach highlights a significant discrepancy in the way the event was narrated in comparison to the media and the government.

These different perspectives indicate a distance between immediate public interest and government official communications regarding the most important sporting event in the world. Girginov (2018, p. 2) argued that 'legacy debates need to consider the relationship between local tangible legacies and the global intangible legacies'. However, what the analysis has shown is that what reverberated in the British national and local media was the local, and more specifically, material and tangible legacy. Despite the billion-pound investment in the International Inspiration Programme, the local negative results received more attention than the efforts made abroad. Legacy, in this sense, despite its multiple meanings and interpretations, has a strong root in the concrete and immediate dimensions. What attracted the attention of the media in the first place was the venue that could be seen, touched and visited by local citizens after the Games: the Olympic Stadium.

It is possible to say that, based on the analysis in this thesis, the media was responsive to the original promises and focused on monitoring their delivery. However, the struggles of East London residents did not receive much attention in the context of the legacy. Although regeneration of Stratford was a promise, the articles making up the British News Media corpus lacked an explanation of what regeneration meant and what the cost of this model of development for the local communities was.

Another point to be highlighted in this analysis is the influence of the past in the discussion of the present. As the creation of white elephants is well-known in the Olympics history (Leopkey & Parent, 2012), the attention paid to places like the Olympic Stadium was

expected. Past negative experiences set the agenda of the media once the host city was 'awarded' the title. These expectations reverberated in the legacy narrative and became central in the discussion.

Regarding the public figures connected with London's legacy, the Olympics has proven to be a fertile soil for the development of their careers. Boris Johnson is the most prominent example, as his intervention in the management of the Olympic Stadium was considered successful. The negative legacies of his government are also widely documented in the failure to deliver a tangible legacy as a result of the International Inspiration Programme. This environment of the media, where politicians tend to make their contributions public, is used to dispute memories once an event becomes historical. This is illustrated by Boris Johnson's attempts to trigger a positive memory of joy and celebration provided by the Games' spirit in 2012, and to connect his image to the fact that he was responsible for that emotion. In this context, the narratives of the media coverage of London's legacy are disputed not just in terms of authorship, by which I mean the actor's responsibility for the successful delivery of the material legacy. The legacy is also disputed in the field of emotions and sentiment, as a powerful nostalgic element that supported a political career that, especially after the Brexit referendum, needed a strong nationalist discourse of pride and a vision for the future.

4.14 Contrasting media outlets

Myles (2010) highlighted how media discourses represent neo-liberal interests in stories of urban regeneration:

> Since the 1990s in the UK the role of the media and journalism has been enhanced in the process of planning and urban regeneration. City press journalism has moved beyond straightforward reporting, or comment, on these issues to a much greater involvement in them. (Myles, 2010, p. 56)

The author argued that media companies have an interest in the economic growth of the cities where they operate, as the urban development results in financial benefits for themselves.

Given the importance of understanding how such interests shape media narratives, this section will look at the differences in the way the four main news outlets selected for qualitative analysis approached Olympic legacy in their reports.

4.14.1 The companies

To provide a brief overview of these outlets, the BBC (British Broadcasting Corporation) is the national broadcaster of the United Kingdom. According to the company's website, the BBC is primarily funded by the licence fee, which is paid by customers to get access to their content, and supplemented by income from commercial subsidiaries (BBC, 2022).

The Guardian is a daily British newspaper whose funding is managed by the Scott Trust, founded by John Russell Scott's family fortune donation in 1936. The organisation does not have a proprietor. Being not-for-profit, the money raised by donations or content funding is reinvested in the company's journalistic production. The organisation defines itself as driven by liberal and progressive values (The Guardian, 2016).

The Daily Mail is a daily British tabloid newspaper. Unlike the BBC and The Guardian, it does not have a public statement on its websites revealing its funding sources. It is owned by Jonathan Harmsworth, who inherited the company founded by his great-grandfather, Harold Sidney Harmsworth. A self-declaration of the Daily Mail's political alignment was not found. However, a survey published by YouGov (Smith, 2017) showed that the newspaper was considered by respondents as the most right-wing among eight options (44% said it is very right-wing and 26% fairly right-wing).

The Daily Telegraph is a British newspaper owned by Frederick Barclay. It was bought by the billionaire twin brothers David Barclay and Frederick Barclay in 2004. Although The Telegraph's funding is also not explicitly publicised, the newspaper charges for access under

a paywall. The Telegraph openly supported Boris Johnson's candidature as Prime Minister in 2009 (Telegraph, 2019).

In a comparison of impartiality among seven news websites by Ofcom News Consumption Survey (Ofcom, 2022), the BBC was considered the second most impartial source of information by respondents (64%), behind Sky News (65%). The Guardian had 54% of the interviewees saying it was impartial and Daily Mail, 35%. The Telegraph is not listed in the survey results.

4.14.2 The coverage

By looking at the most frequently occurring words in every single newspaper in the British News Media Corpus and contrasting the quantitative data with a qualitative close reading of the articles, it is possible to see which topics in relation to legacy were prioritised by the different publishers. The BBC presented a wide range of distinct topics in the legacy narrative, with a distribution between the construction of the park, the access of young people to sports, physical education provision in schools, east London redevelopment and political actions, with the latter mostly focused on Boris Johnson and Sebastian Coe.

The Guardian focused its coverage on the politics and economy of the legacy, with the cost of the Games and the regeneration of East London being two of the main topics highlighted in their articles. 'Public' is among the most frequently occurring words in The Guardian, often associated with public money, public spaces, public interest, public use and public spending. This 'public' character of the coverage is clearly illustrated in 'The Olympic legacy will be a festival of the people' (Hone, The Guardian, 2012). The newspaper's website also gave considerable space to discussions around the Olympic Stadium although it presented a balanced approach in comparison to the other topics it covered.

The Daily Mail was mainly focused on the discussions around the destiny of the Olympic Stadium, with 'Stadium', 'West' and 'Ham' being the most frequently occurring words. Other words associated with the story – like 'athletics' and 'Tottenham' – also appeared among the

top 30 most mentioned. There is an imbalance in the Daily Mail coverage, which prioritised discussion of the Stadium over other topics that received less attention, such as school sports activities.

For The Telegraph, the main discussion was developed around 'school sport'. Among the main topics are 'schools', 'funding', 'participation' and 'children', all of them highly associated with the school sports provision. In this cluster, there is a change over time in the importance of 'school'. In 2010 and 2011, 'West Ham' was mentioned three times more than 'school'. From 2012 onwards, 'school' gained more space in the coverage. The chart below illustrates the variation over time. The chart shows in blue mentions of the word *school* and in orange of the word *ham*, both by The Telegraph, over the years.

Figure 37 – Temporal distribution of the word 'school' and 'ham' in articles published by the Telegraph (based on word count per year)
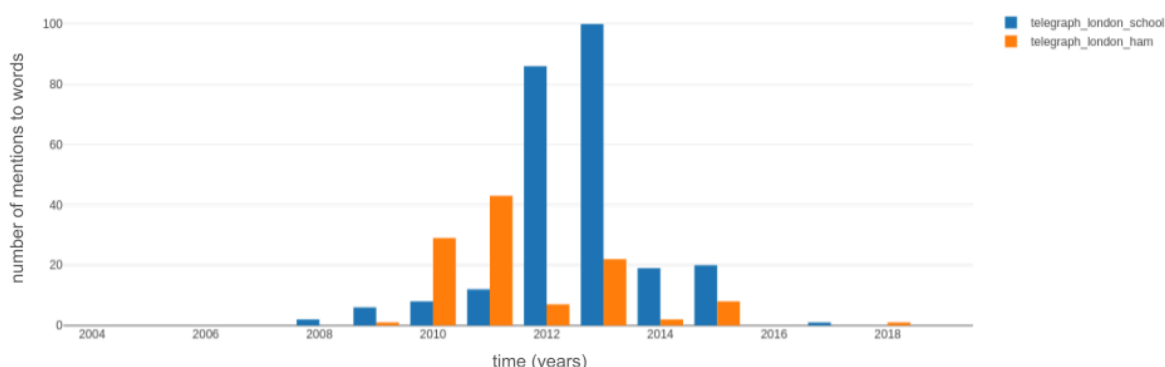
Although, as mentioned above, The Telegraph made public its preference for the Conservative party and the figure of Boris Johnson in particular, the coverage of the discussion around schools was very critical. The government funding cuts were denounced in their publications and most of the articles on the topic were demanding action for the delivery of the 'inspiration' promise, especially after 2012.

For this analysis, the political alignment of news outlets was not used to categorise the articles studied or to guide the interpretative work. While it is useful to take into account the potential biases and interests of specific media organisations, this section also highlights how

newsrooms, although highly guided by their editorial boards, are also a place of dispute, where journalists work on and sometimes beyond the margins. The fact that The Guardian gave space for the Triathlon Homes director to write about the benefits his company had brought to East London may frustrate some expectations regarding the outlet's editorial line. On the other hand, the 'activism' of The Telegraph for better sport infrastructure at schools may also astonish a researcher excessively guided by political alignment biases. Moreover, this study's approach across a range of outlets, rather than focusing on the specific biases of a single outlet, can provide a more holistic understanding of national media discourses.

4.15 Conclusion

This chapter discussed the results of the experiments performed with the dataset composed of news articles published by the British media on the legacy of the 2012 London Olympics. Aspects of the media coverage were explored using NLP techniques, network analysis and discourse analysis.

This chapter covered the disputes over the use of the Olympic stadium between West Ham and Tottenham and the critical and pessimistic tone of the coverage before the event from the news media regarding London's capacity to deliver the plan. This was particularly salient in the news media response to the legacy plans with regards to the projects for young people and lack of infrastructure for sports in British schools.

To understand the discourses of public officials, I analysed the government's framing of legacy in texts published on their official websites. The chapter shows that there was a cyclical pattern followed by the legacy coverage, in which journalists made use of past examples to project possible futures, as evidenced by the network analysis. The analysis highlighted the role of important political figures, such as Boris Johnson, in the legacy discourse and how his personal political career benefited from the Olympic legacy.

Finally, the media silences and the activists' warnings regarding the negative impacts of hosting the event were identified, with a particular focus on the lack of new jobs and

affordable homes in East London. With regards to the International Inspiration Programme, this chapter highlighted that there is a conflict between the project of inspiring people around the globe that was used as the main legacy discourse in government communication, and the news media's interest in national and local issues.

At the end, a discussion on the news outlets' economic organisation and the characteristics of their coverage was also provided. The BBC presented the widest diversity of topics in the coverage of London's legacy, including the Olympic park construction, the access of young people to sports, physical education provision in schools, east London redevelopment and political actions, with the latter mostly focused on Boris Johnson and Sebastian Coe. The Guardian focused its coverage mainly on the politics and economy of legacy. It focused on the discussions around public spaces and public interest. The Daily Mail, however, dedicated most of its coverage to the dispute about the future of the Olympic stadium. The Telegraph focused on education and sports, especially on the lack of infrastructure and the promises of increasing participation of young people in sporting activities.

The analysis presented in this chapter revealed an extensive focus of the British media coverage of London 2012 Olympic legacy on tangible legacies, in particular the afteruse of the Olympic Stadium and the projects to expand access to sports activities for young people in British schools. The media framing differs from the discourses of government websites, which focused primarily on the International Inspiration Programme, an example of intangible legacy. Moreover, I identified how past examples of legacies – in particular the negative legacies – are brought up by the media to compare London with previous host cities, but also to alert future concerns with regards to the delivery of positive and long-lasting legacy. An important finding concerns the figure of Boris Johnson as a central figure responsible for shaping legacy narratives. The Mayor of London embodied legacy as he centralises the discussion on his figure, which despite the risks, seem to have contributed to producing positive narratives in the media with regards to his efficiency as a politician. The implications of these findings, how they compare with Rio 2016, and how they contribute to the wider understanding of how legacy is narrated in the media will be provided in Chapter 7.

In the next chapter the news articles published by the Brazilian media on the legacy of the Rio 2016 Olympic Games will be analysed and the results will be discussed.

Chapter 5: The Brazilian media narrative of Rio's legacy

In this chapter, coverage of the Rio 2016 Olympic legacy by the Brazilian media will be analysed. The same methods applied to the three English corpora[64] were replicated for the texts written in Brazilian Portuguese. However, the challenges of multilingual analysis using digital methods will be discussed first, as they impacted the findings. In the next section, I will cover the main barriers encountered while studying Brazilian content and how the issues were overcome.

5.1 Data scarcity, algorithms and translation

As discussed in Chapter 3, there was a gap between the amount of content produced by the British news media in comparison to Brazil. One of the potential reasons for this discrepancy is the organisational structure of the Brazilian press in comparison to the United Kingdom. Brazil appears in 111th position in the ranking of the 2021 World Press Freedom Index. One of the reasons for the negative score is, according to the Reporters Without Borders' document (Reporters without Borders, 2023), the concentration of media ownership in the hands of business families that have a close relationship with the political class. The lack of a strong public broadcast organisation - like the BBC - combined with the oligopoly of the media companies resulted in less diversity of publishers and, therefore, less information availability for researchers interested in multiple sources.

Another limitation concerns the quality of the networks produced. The outputs from Spacy in Portuguese for entity recognition presented a much poorer result in comparison to those produced in English. In the process of disambiguating the entities by combining Spacy outputs with Wikifier, many entities were misread. To avoid mistakes in the analysis, entities

---

[64] British News Media Corpus, British Government Corpus and British Activist Corpus.

found that did not correspond to the 'vocabulary'[65] of the Rio Olympics were manually identified and then a search in the corpus was performed to find the reason why the entity was being misinterpreted by the algorithms.

This approach to manually identifying mistakes in entity recognition was the result of a qualitative intervention in the quantitative analysis. Expert knowledge on the topic was essential for these mistakes to be recognised and for developing an approach to overcome them. In the network, for example, one of the central entities is CPLP Games. This entity refers to the Wikipedia page CPLP Games (Wikipedia, 2023), which describes the topic as 'a multinational multi-sport event organised by the Community of Portuguese Language Countries (CPLP)'. However, when looking closer at the corpus, it was verified that CPLP Games was attributed to the word *jogos* – Games, in Portuguese. When the word *jogos* appeared alone, it was wrongly attached to CLPL Games. When *jogos* was mentioned together with another word such as in *jogos olimpicos* (Olympic Games), it was correctly attributed to Olympic Games. Entities were also attributed to wrong Wikidata entries that were still among the Olympics vocabulary. An example that was found was *jogos olimpicos* being attributed to the 2012 Summer Olympics – the London Games. *Velodromo* was attributed to 'Velodromo Paulistano' – which is located in Sao Paulo, not in Rio; Olympic Stadium – or *estadio olimpico* – was wrongly associated with 'Estadio Olimpico Monumental', which is based in Porto Alegre, south Brazil, again not in Rio. These mistakes made the analysis of entities in the networks impracticable for the aim of identifying with precision specific entities and their connections. A general view of the network was still considered valuable, however, bearing the limitations of Wikidata entity mismatch in mind. The challenges of applying algorithms of SA to the Portuguese corpus will be addressed in more detail in the following chapters.

---

[65] By the 'vocabulary' of the Rio Olympics, I mean the set of words, organisations, entities and government bodies that are mentioned in the media coverage.

Unveiling these limitations is important for reflecting on the challenges of working with multilingual analysis. These limitations go beyond language, and point to what Sibilia (2008) called a techno-apartheid, referring to the inequalities in terms of access and technological production in the Global North and South. Bon (2021) referred to this phenomenon as 'digital coloniality' and pointed out the importance of making people from these regions co-researchers, co-creators, and co-designers of technologies. Making use of technology developed by and for the Global North, or even more specifically the Anglosphere, limits our mechanisms to study social phenomena, generating geographical boundaries in the online territory.

A final important topic to be mentioned is how the challenges of translation were addressed in this chapter. Acronyms for Brazilian government bodies will be explained in English but kept in their original Portuguese form. There was a particular challenge in translating names of public and government institutions as they did not necessarily have a correspondent in the United Kingdom. For example, there was a strong emphasis on legal debates in the Brazilian press coverage of Rio 2016 and MPF (*Ministerio Publico Federal)* was one of the most frequently occurring bigrams that will be discussed later in this chapter. Its closest 'equivalent' in the United Kingdom is the Crown Prosecution Service. As the Brazilian system does not work as a representative of the Crown, the term would be inappropriate. Having said this, the Brazilian Public Prosecution Service will be referred to using its Portuguese acronym: MPF. The same approach will be applied to the other bodies to avoid inconsistencies.

In this chapter, the analysed data is the Brazilian News Media Corpus. The sources of news articles used for the qualitative analysis, present in this corpus, are Globo (from Globo Group: includes G1, O Globo, Globo Esporte, Extra, TV Globo, SportTV, Globo News, CBN Rádio, Valor Econômico and Revista Época), Folha (from Folha Group: includes Folha de São Paulo online, UOL and Band) and Estadão (from Estado Group: includes O Estado de São Paulo online).

To guide the qualitative exploration, the bigrams produced with Python scripts – presented in Chapter 3 – were used. Voyant Tools was used to access specific sentences and contexts, replacing TIME, due to the issues accessing TIME at the time of analysis, as described in section 3.5.1.

5.2 The beginning: 'A fantastic excuse to change Rio'

In 2011, two years after Rio was appointed the next host city of the Summer Olympic Games, Nick Clegg, Deputy Prime Minister of the United Kingdom, and Jeremy Hunt visited Brazil for the conference: 'From London to Rio - an Olympic legacy', as recorded in the article entitled 'Rio's Olympic budget has to be defined now, says British Minister' (Globo, 2011). According to the article published by Globo, Jeremy Hunt recommended that Brazil should start planning its budget for the 2016 Olympics in advance, emphasising that 'it has to be now'. Hunt's advice was based on London's previous experience, so he argued, of not having prepared the budget from the beginning, and thus having to revise their planning in 2007. During the event's opening ceremony, Rio's Mayor Eduardo Paes said he believed in the Olympics as a 'catalyst element to change Rio', highlighting the importance of having a material legacy that could be reached through urban transformation. This article illustrates fundamental points that will be discussed in detail later: the importance of budget planning in advance to guarantee the success of the event and the Mayor's understanding of legacy as a synonym for urban regeneration.

Although the meeting was presented as an opportunity for Rio to learn from previous experiences - such as London's - Rio already had a background in mega sporting events, having hosted the *Jogos Pan-americanos* in 2007. On that occasion, as reported by Veja (Lemos, Veja, 2011), the initial budget to construct the Engenhao Olympic Stadium of R$60 million (£9 million) ended up costing 6 times more, R$380 million (£58 million).[66] Henrique

---

[66] Exchange rate of £1 = R$6.19 consulted on the 18th of August 2022.

Meirelles, ex-president of the Brazilian Central Bank and member of the Olympic Public Council, said that this mistake would not happen again.

Rio's Mayor's discourse was also 'optimistic', as in the article published by Globo in March 2012 entitled 'The Olympics are a fantastic excuse to change Rio, says Mayor' (Globo, 2012b). According to the publication, Paes referred to the arguments that 'changes could occur despite the Olympics' as 'idiotic'. For him, it was necessary to have the event as a 'challenge' for the transformations to occur. When asked what the legacy of the Rio Olympics in 10 or 20 years would be, Paes responded by mentioning only the planned improvements in transportation to facilitate, in his words, the access of the 'poor [living in the periphery] that need to go to work [in the city centre]'. To the question of whether Rio would create an institution to manage the legacy as London had created the Legacy Corporation, the Mayor said he would not create 'another space for corruption'. The Mayor seemed to see the idea of sharing power with other government bodies as challenging and, therefore, he advocated for his control over the plans instead of the creation of a specific institution.

5.3 Ministry, authorities and the justice system: 'less democracy is sometimes better'

The sentence 'less democracy is sometimes better' was uttered by Jerome Valcke, FIFA chief, in 2013. A year before the 2014 FIFA World Cup in Brazil, Valcke said to the BBC that the structure of the Brazilian State, divided into various levels of government, caused difficulties for delivering a major sporting event. In the article, Jerome Valcke said that:

> When you have a very strong head of state who can decide, as maybe [President Vladimir] Putin can do in 2018 ... that is easier for us organisers than a country such as Germany, where you have to negotiate at different levels. The main fight we have [is] when we enter a country where the political structure is divided, as it is in Brazil, into three levels - the federal level, the state level and the city level. [There are] different people, different movements, different interests and it's quite difficult to organise a World Cup in such conditions. (BBC, 2013).

The FIFA chief's preference for authoritarian regimes may reflect the complex governance apparatus of government bodies and independent institutions responsible for supervising the implementation of public policies in Brazil. Among the 15 most frequently occurring bigrams in the Brazilian News Media Corpus are *Ministerio do Esporte* (Ministry of Sports), *Autoridade de Governanca* (Rio 2016 Governance Authority - AGLO), *Governo Federal* (Federal Government), and *Ministerio Publico* (Brazilian Federal Public Prosecution Service - MPF). The presence of government bodies in the list of bigrams analysed in Chapter 4, from the British coverage of London 2012 Olympics, was much less significant, with the 'delivery authority' being only the 27th most frequent bigram. In contrast, the Brazilian AGLO is the 4th most frequently occurring bigram in the Brazilian News Media Corpus. Although this finding does not necessarily point to a country with more democracy, as argued by Valcke, it shows how strong the presence of the State, and specially the justice system, was in the media narrative. A background analysis of the Brazilian political scenario since 2013 is thus necessary to understand the context in which the media coverage of Rio 2016 occurred.

Andrew Zimbalist (2017) presents an overview of the social, political and economic situation in which the event took place. Among the topics pointed out by Zimbalist, such as the economic recession, the increase of violence and president Rousseff's impeachment, one is particularly relevant to the discussion here as it helps to explain why the bigrams mentioned above have such relevance in the media narrative: The Car Wash operation.

In 2014, the Brazilian Federal Police started the Car Wash operation aiming to dismantle corruption schemes in the Brazilian government. President Dilma Rousseff's party, the Workers Party (PT), was highly affected by the operation and some of its members were arrested. Among them was former President Luis Inacio Lula da Silva, arrested in 2018.[67]

---

[67] All charges against the ex-President were cancelled by the Supreme Court in 2021. The judge responsible for Lula's sentence, Sergio Moro, was considered partial, having his personal involvement in the investigation detailed in leaked conversations on Telegram. Messages were published by The Intercept in the scandal known as *Vaza Jato*.

Several studies on the media coverage of the operation (Cioccari, 2015; Fontes, Ferracioli & Sampaio, 2016; Gomes & Medeiros, 2018; Fontana, 2020) have shown how the discourse of corruption was framed as intrinsic to Brazilian politics. Furthermore, the city of Rio de Janeiro in particular had been impacted by corruption schemes in the years before and after the Olympics, with its last four governors[68] (from 1999 to 2018) having been arrested. As a consequence of the Car Wash operation, corruption schemes became one of the main topics of Brazilian news media, and this reverberated through the coverage of the Olympic Games. As previously discussed (see Chapter 2) the legacy timeline often comprises significant political shifts from the candidacy to host the Games to the delivery of the event itself.

This background helps to understand some of the most mentioned words in the corpus: 'millions', 'federal', 'government', 'construction', 'arenas', 'city hall', 'facilities', and 'authority'. They all refer to the relationship between the different levels of governance in the Brazilian State and the economic investment in the facilities constructed for the Games - especially sports venues. The responsibility for funding and delivering these facilities, however, was disputed by the different levels of the Brazilian government. This is evidenced by the salience of the bigram 'Ministry Sport' in the corpus, which relates to the dispute between the Municipal and the Federal Governments. According to the article 'Mayor office launches legacy plan and exposes battle with Ministry of Sports' (Filipo, Globo, 2015), the Mayor of Rio's office launched their own independent legacy plan without the participation of the Federal Government. Pedro Paulo Carvalho, the Mayor's secretary, announced Rio's plan based on a list of partnerships with the private sector to deliver 10 venues for the Games [69] and one outdoor area.[70] Rio's Mayor accused the Federal Government of abandoning the city, leaving the construction of the Olympic Park to be managed by the Mayor's office, as

---

[68] Anthony Garotinho (1999-2002), Rosinha Garotinho (2003-2006), Sergio Cabral (2007-2014) and Luiz Fernando Pezão (2014-2018).

[69] Arena Carioca 1, Arena Carioca 2, Arena Carioca 3, Velódromo, Centro de Ténis, Arena Rio, Pista de Atletismo, Arena do Futuro, Estádio Aquático and Parque Radical de Deodoro.

[70] Via Olimpica. The area would become a public park.

published by Globo in '14 months to Rio 2016, uncertainty over main legacy's future*
(Guerra, Globo, 2015).

5.4 Promising Games, challenging reality: the pre-Olympic discourses

'School named after Socrates will be part of Olympic legacy' (Globo, 2012c) and 'After
Games, handball arena will be transformed into four schools' (Estadao, 2013) are examples
of news articles published by the press on the promising legacy of the Rio Olympic Games.
The first, published by Globo in 2012, narrates the inauguration of the Municipal School
Doctor Socrates, in honour of the ex-Brazilian football player. The School, according to the
article, was part of a programme entitled Experimental Olympic Gymnasium (GEO). The
programme's objective was promoting high-performance sports for children.[71] Official
documentation of this project, however, is not available. As Rio did not publish a plan for its
Olympic legacy, information is scarce, and can only be found on the government's website.[72]
According to the rededoesporte.gov.br, the programme involved the construction of five
schools, focused on preparing the athletes of the future. While Globo's article presents an
accomplished legacy four years before the Olympics, the second article, published by Estadao
in 2013, is about the future: the temporary facilities constructed to host the handball
competitions would be used to build four other schools after the Olympics.[73]

Both articles illustrate how optimistic discourses appeared in the news media in the years
before the Olympics, often strongly associated with the figure of the Mayor making promises

---

[71] This differs from London's idea of including young people in the legacy plan. London's legacy plan
was more focused on expanding opportunities for young people to practice sports in general. Rio's is
focused on developing the athletes of tomorrow.

[72] Available at
http://rededoesporte.gov.br/pt-br/megaeventos/paraolimpiadas/legado/ginasio-experimental-olimpico.
Accessed on 12th June 2022. Archived at
https://web.archive.org/web/20230612101405/http://rededoesporte.gov.br/pt-br/megaeventos/paraolim
piadas/legado/ginasio-experimental-olimpico.

[73] Those four schools are part of a separate project. They are not included in the GEO.

or inaugurating facilities that were 'included' as legacies of the Olympics, even though they had not been presented as an official project in advance. In another article published by Estadao, entitled 'Olympic dream changes Rio's landscape' (Pereira, Estadao, 2014), the construction of the Olympic Park is described as 'frenetic', and the construction of tunnels, metro lines, and BRT (Bus Rapid Transport) is referred to as 'unprecedented' by the journalist, Pablo Pereira. The investments were at the time around R$ 37 billion. Paes is quoted comparing the amount of money invested in Rio with Beijing and London: 'Beijing, in 2008, spent R$ 65 billion; London, in 2012, spent R$ 41 billion, with R$ 33 billion paid by the government'. The Mayor added that 57% of the investments in Rio came from Public-Private Partnerships (PPP), with public money constituting just 43% of the total spent. In another example 'Rio focuses on efficiency for the Olympic Games' (Estadao, 2015), the director of the Municipal Olympic Company (EOM) - responsible for managing the construction of stadiums and facilities for the Games - said that while cities such as Beijing had presented luxurious stadiums, like the Bird's Nest, Rio would be remembered for its efficiency, leaving for its citizens a 'tangible' legacy. According to the article, the organisers' main concern was 'avoiding white elephants'.

In parallel with the articles on positive expectations for the Olympic legacy, others were published describing concerns about the future, for example 'Problems in construction threaten city's Olympic legacy' (Estadao, 2016). The article enumerates several failures in the facilities constructed as part of the Olympic 'legacy',[74] such as the inefficiency of the BRT lines (overcrowded and precarious) and the delay in delivering Line 4 of Rio's Metro, which would connect the city centre to the Olympic Park. The text also mentions the collapse of a bike path constructed as part of Rio's urban regeneration for the Olympics.

---

[74] Legacy is used in quotes because it is inappropriate to refer to a legacy plan, as it did not exist as a document until then.

5.5 Guanabara Bay and the 'green' Games slogan

In *'Lars expects Olympic legacy near zero and defends Sailing in Buzios'* (Globo, 2014) the focus is on the pollution of the Guanabara Bay. Lars Grael, former sailor, was interviewed and said he believed in decontamination before 2016. 'Guanabara Bay' is among the top bigrams (23rd) mentioned in Rio's corpus, occurring 80 times. The article *'Sanitation becomes an Olympic challenge because of the Guanabara Bay'* (Globo, 2013b) reported on the lack of sewage systems in most cities around Rio, whose waste was dumped into the rivers. This made Guanabara Bay a problem for the Olympics as it was intended to be the future venue for Sailing competitions. A project to clean up the Bay was among the central projects of the Games. The problem, however, was not solved before the Games were hosted, as reported by Globo (Bortoloti, Globo, 2016). The solution found was to install 'nets' to at least prevent large waste from being seen in the Guanabara Bay during the competition, although the athletes did have to compete in polluted water.

The discussion around Guanabara Bay is part of a broader topic in the Rio Olympics narrative. In Rio's opening ceremony, sustainability was one of the main themes. This is a key difference from the London 2012 Games, which was more focused on the pride of the industrial revolution. A big part of the opening ceremony was devoted to informing the audience about the importance of preserving nature, to ensure the long-term existence of the Games. This was evidenced by the Olympic rings formed by trees and a short film about global warming, exhibited during the ceremony, showing how cities would be flooded if the sea levels continued to rise (Olympic Channel, 2016).[75] The fake trees in the Olympic rings received real seeds from the athletes, which would go on to form, according to the ceremony, the Athletes' Forest in Deodoro Park,[76] another legacy of the Games. The Forest, initially planned to be planted in 2017, actually materialised only in December 2019. According to Globo (Azevedo, Globo, 2021) the reason for the delay was that the promise was only

---

[75] Available at https://olympics.com/en/video/opening-ceremony-rio-2016-olympic-games. Accessed on 15 June 2022.

[76] Deodoro Park hosted rugby sevens, shooting, hockey and equestrianism among others.

'orally/informally made', without a signed contract. Because of that, the local organisers had no obligation or funding to fulfil the promise. In 2019, a solution was found: companies in green debt - for damaging the environment - committed to paying for the forest to be planted.

The attempt to make Rio the 'Green Olympics' was, however, frustrated as the opening ceremony theme was not sustained by a legacy promise published in the form of a plan. If one searches in 2022 on Google in English for 'Rio Olympics green', the results contain mostly articles in the international press questioning why the diving pool suddenly became green during the Games, for example 'Why is the Olympic diving pool green? The good news is it's not urine' (Hunt, The Guardian, 2016). This problem, connected to an overabundance of algae, attracted more attention than the 'green' message of the opening ceremony.

Due to the lack of a public document describing the plans for Rio's legacy, in July 2016 Globo radio CBN published the article 'Environmental legacy: the great debt of Rio Olympics' (Soares, Globo, 2016). The author Lucas Soares revisited Rio's dossier, written for its candidature in 2009, where 49 promises were made, of which only 16 were entirely fulfilled. For the journalist, the environmental legacy was the most neglected, and he highlighted the most prominent failures: the lack of planted trees, the failed decontamination of Guanabara Bay and the unreached goal of neutrality of carbon emissions. In the article, the response from the Public Olympic Authority (APO) was that the dossier was conceptual, and not everything described there was to be executed, 'only what the IOC considered as essential'.

5.6 The Tennis Centre, Velodrome and fire: the fall of the material legacy

Among the most frequently occurring bigrams in the Brazilian News Media Corpus are *Centro tenis* (Tennis Centre), with 149 mentions, *Arena Carioca* (Carioca Arena), 125, *instalações esportivas* (sporting facilities), 74, *instalações olimpicas* (Olympic facilities), 72, and *Velodromo parque* (Velodrome Park), 57. Three of these bigrams refer to sporting venues

constructed for the Games.[77] Also among the bigrams and connected to the previous words is *corpo bombeiros* (fire brigade), mentioned 67 times. On 30 July 2017, Globo radio CBN published the news article 'Fire breaks out in Rio de Janeiro Velodrome' (Globo, 2017). The fire was caused by a hot air balloon crash landing and mainly affected the roof of the venue. The quality of the track was threatened as it relied on 24-hour air-conditioning to keep the right temperature for the siberian wood, which cost approximately R\$ 11 million a year in electricity bills (Globo, 2017). But why are 'fire brigades' mentioned more than 60 times in the Brazilian News Media Corpus? The answer is in 'New fire breaks out on the roof of the Rio Olympic Park Velodrome*'* (Gaier, Globo, 2017). After the first fire in July, there was a new fire in the same venue in November 2017. The reason was the same: a hot air balloon.

The fires in the Velodrome are symbolically meaningful in the context of the material legacy of the Rio Olympics, which was strongly focused on the sporting facilities. If the materiality of the legacy was the main focus of the government and the media, the high susceptibility to fires transformed those structures into a symbol of the government's inability to deliver a positive legacy. The Tennis Centre appeared in the news mostly due to lack of maintenance, infiltrations, destruction of its seats and flooding in the years after the Games (Globo, 2021). In 'Rio suffers from expensive and empty arenas after the Olympics' (Estadao, 2017b), published in August, the lack of a permanent use for the Arenas Carioca 1 and 2 is mentioned as an economic problem for the government as their cost was approximately R\$ 45 million that year. The plan for the Arena Cariocas, according to the Mayor, was that Arena Carioca 3 would become a school, Arena Carioca 2 would be transformed into an Olympic training centre for athletes, and Arena Carioca 1 would be handed over to the private sector (Gaier, UOL, 2016). As we will see in the next section, privatising public infrastructure is a common-place in the legacy discourses.

The article also mentions the Arena do Futuro (Future Arena), a pre-fabricated structure that would be transformed after the event into a new building. The place was originally intended to become a school, but a year after the Games, the venue remained in the same place and

---

[77] *Centro tenis*, *Arena Carioca* and *Velodromo parque*.

empty. The reason for that, according to Marcelo Crivella, Rio's Mayor from 2017 to 2020, was the lack of financial resources. The empty or underused venues were many times promised to be privatised, an attempt widely documented by the media coverage.

5.7 Privatising: an attempt to rescuing the legacy

Iniciativa privada (private sector), mentioned 88 times in the corpus (21st in ranking), is among the most frequently occurring bigrams. In 2014, in the previously mentioned article 'Olympic dream changes Rio's landscape' (Pereira, Estadao, 2014), Rio's Mayor, Eduardo Paes, proudly communicated the 'small' percentage of public money invested in the event thanks to the Public-Private Partnerships (PPP). The privatisation of the Games facilities was a discourse strongly shaped by the social, political and economic context in which the event occurred. As discussed above, the Games were surrounded by corruption scandals investigated in the Car Wash operation. President Dilma Rousseff was impeached and accused of manipulating the country's public funds in her favour.[78] The media coverage of the investigation was highly sensationalist, especially on TV (Azevedo Dias, 2018), producing an effect that was referred to by scholars as the 'criminalisation of politics' (Gandin, 2019). The term refers to the idea that politicians are inevitably corrupted, and the private sector is the ideal model of efficient management. As shown by Figueiredo & Amorim (2020, p. 76), this phenomenon started in the 2014 FIFA World Cup, also hosted in Brazil, with the media having dedicated considerable space to the Car Wash operation's denunciation of corruption schemes involving the construction of several venues across the country.

Taking this political scenario as background, the analysis of the texts mentioning *iniciativa privada* reveals the term as first mentioned in the context of a solution to avoid the 'inefficiency' and 'corruption' of the public sector. Later, however, the term is mostly mentioned as a frustrated plan, due to the lack of investors willing to manage the facilities. Rio's legacy plan - published in 2017 - included the 'market's aversion to association with the

---

[78] President Rousseff was later absolved of all accusations.

public sector' in the list of weaknesses and threats to the legacy. The lack of interest of investors in managing facilities forced the government to take control over the maintenance expenses, distributing the management costs among different government bodies (the Federal Government, the Municipal Government and the Army). In January 2018, the article published by Estadao entitled 'Union [Federal Government] will spend R$ 35 million to keep the Olympic Park functioning', mentions the lack of investors, which was recognised by AGLO's President Paulo Márcio Dias Mello. For Dias Mello, the risky investment could be of no interest to the private sector.

The article 'After four months, Olympic Park in Rio has signs of abandonment' (Brito, Globo, 2017) narrates the struggle of Rio's Mayor to find investors. It starts by describing the amount of money (R$ 3.324.576,31) paid 'urgently' to the company *M Rocha Engenharia* to manage the Olympic Park. Despite the investment, the Park looked abandoned. The swimming pool was described as full of algae and mosquitoes. Dangerous electrical equipment was poorly protected with a grid, presenting a high risk of accidents. Art pieces designed by important Brazilian artists like Adriana Varejao were partially destroyed. Eduardo Paes' press officer said to Globo that the city hall had issued two public tenders in which there were no interested parties. The author of the article, Carlos Brito, alluded to two dystopian films, The Omega Man (1971) and I Am Legend (2007), to describe the scene: 'At first, the landscape looks like the scenarios of The Omega Man and I Am Legend – science fiction plots that show empty cities after a cataclysmic event', wrote Brito. The dystopian discourse of the legacy (see Chapter 2.4), was directly referenced here and resonates with Zimbalist's (2017) reference to the Rio Olympics as 'hell'.

Finally, in 2021, Globo published 'Olympic legacy: Rio 2016 Tennis Centre goes to the private sector' (Globo, 2021b). According to the article, Rio's City Hall had finally started the process to hand over the Tennis Centre and Arenas Carioca 1 and 2 to the private sector. The project resumed in 2021 as Eduardo Paes had been elected Mayor of Rio again in 2020. In the article, the Mayor is quoted saying that he was now deploying the initial project of the Games: 'It was one of the cheapest Olympics in history, with the highest volume of private

investment and there isn't a white elephant here. The urban legacy is also there'. The Mayor focused on the pillars already discussed in this thesis: the amount of public resources invested, the work of private investment - as it was free of public money - and a fear of white elephants. Paes is a central figure in the coverage of the Rio Olympics and the narratives surrounding the politician will be analysed in the following section.

5.8 Eduardo Paes: the Mayor's trajectory

Rio de Janeiro was elected the host city of the 2016 Summer Olympic Games in 2007. In 2008, Eduardo Paes was elected Mayor of Rio. Paes governed the city for four years (2009 - 2012), and was reelected in 2012 for a further four years (2013-2016). Unlike Boris Johnson, elected three years after the Olympic planning started for London 2012, Paes was involved from the beginning until the year of the Rio Olympics as Mayor. In the Brazilian News Media Corpus, the news articles covering the Olympic legacy of Rio began to be published by the Brazilian media in 2011, with the previously mentioned Globo article reporting on the visit of London's delegation to Rio. From this date, Eduardo Paes is cited as the main figure behind the Games, with his name being the 6th most frequently occuring bigram in the Brazilian News Media Corpus (161 times).

Due to the lack of planning documentation and, initially, of a governance authority to manage the legacy, Paes was cited in the news articles as the main spokesperson, presenting him as a kind of owner of the Olympics. In 'Eduardo Paes does not reveal Olympic budget and shoots against FIFA' (Coelho, Globo, 2013), the Mayor argued that unlike the World Cup, Rio had proper planning. However, every Olympic legacy facility to be constructed for the Games was referred to as a 'promise'. This idea of promise carries a sense of a personal rather than institutional commitment in relation to the legacy. Paes, in this sense, did not represent the City Hall or the figure of the Mayor of Rio de Janeiro. Paes was the City Hall himself and took responsibility over the entire event as a result of his personal commitment. This is evidenced by the fact that in the first paragraph of the news article, the author Henrique Coelho used the expression 'to make a point of', which produces an effect of equivalence

between the oral promise and the expected official documentation: 'Mayor Eduardo Paes did not reveal what the city's budget will be for the 2016 Olympic Games, but made a point of stressing that most of the spending on the Olympics will be made with private resources and public-private partnerships'.

Figure 38 − Network visualisation of entities cited in news articles published by the Brazilian media covering the legacy of Rio 2016 Olympic Games

The network presented above shows the name of Eduardo Paes on the centre left of the image, directly connected to 'CPLP Games' (which refers to the Rio Games), 'Olympics' and '2012 Summer Olympics' (which also refers to the Rio Games). The prominence in the

legacy of the figure of Paes meant he would receive recognition for his positive work but also criticism for what was not achieved as expected. In the article 'Eduardo "Pereira" Paes, the Brazilian Olympics' Haussmann' (Carvalho, Estadao, 2016), the author Leonardo Carvalho compares Paes with two figures, Francisco Pereira Passos and Georges-Eugène Haussmann. The first was Mayor of Rio between 1902 and 1906 and was responsible for a significant transformation of the city, especially affecting urbanisation through processes of gentrification. Pereira Passos, according to the article, had been inspired by Haussmann, responsible for similar changes in 19th century Paris. The model of development adopted by Rio was criticised in the article and referred to as a process of 'sanitising the city' by sending the poor to live out of and further from the central areas. This article by Carvalho is an example of how the legacy narrative was interconnected with conceptions of urban development. In this context, the Olympics was presented as an opportunity for accelerated urban regeneration, but it brings together the logic of social exclusion that affects urban centres globally as stated by Harvey (2012). While the expectation of urban development as a result of the Olympic Games already constitutes the idea of legacy, urban development should not be taken as an objective concept, but rather as one that triggers the question: what sort of the development do we expect? Or, as defended by Harvey, what kind of city do we desire?

Like Boris Johnson, Eduardo Paes was also referred to as a 'legacy' of the Games. The article 'Olympics raises Paes as candidate for Governor of Rio in 2018' (Arbex, Folha, 2016) describes the Mayor's Party meeting to decide on the best strategy to advertise his preferred successor's candidacy, motivated by the euphoria of the Games. According to Folha, even though the Mayor had been booed during the closing ceremony, he was the main beneficiary of the Olympic legacy. The article stresses Paes' intention to be a candidate for Governor of Rio de Janeiro (the State) in 2018 and his Party's belief that his success in managing the Olympics would be enough to secure the election of Pedro Paulo as his successor in 2016. For his Party, the 'legacy' would be powerful enough to hide Pedro Paulo's accusations of assault against his wife. The main strategy would be to present Pedro Paulo as the preferred candidate of the Mayor responsible for delivering the Olympic Games. The article concludes

by presenting the opposition's strategy of highlighting that the investment in the Games had not been enough to leave a legacy.

His Party's strategy failed. Paes' candidate received only 16% of the votes in the election's first round. The winner, Marcelo Crivella, had been a critic of Rio's Olympic legacy, as shown in '"Outside the election, Paes runs to church", says Crivella' (Nogueira, Folha, 2016). In the article, the candidate who would become Mayor that year said that the improvement of public transportation infrastructure came with an increase in the cost of tickets, stressing the reduction of ticket prices as one of the main promises of his candidature. Crivella said to Folha that he wanted 'the Olympic legacy to be really for the people'.

In 2018, Paes also lost the election to become a Governor of the state of Rio de Janeiro. Paes received 40% of the votes and the winner, Wilson Witzel, 59%. Witzel was, however, later impeached in 2020, accused of corruption in the use of public money during the emergency state caused by the Covid-19 pandemic. In 2020, Paes was again elected Mayor of Rio (64% of the votes) against Crivella (35%).

It is not clear that the Olympics brought the same political benefits to Paes as they brought to Boris Johnson's career. The Games euphoria could not be as easily converted into support for the Mayor due to the lack of a recognisable legacy, in parallel with the scandals surrounding his work as Mayor. As documented by Zimbalist, one such scandal was the collapsed bike path in Rio, constructed by the group Concremat, which 'belongs to family members of Rio's tourism secretary, who was also treasurer of both of Paes' campaigns for Mayor' (2017, p.5).

This, however, was not the end of Paes' attempts to create a positive legacy discourse around his name. After becoming Mayor of Rio again in 2021, news about Rio's Olympic legacy – outside of the scope of the Brazilian News Media Corpus – started to be published that not only referred to the past but also now reported on the present and the future. In 'Olympic Legacy: Rio begins work to transform Arena do Futuro into schools' (Buzzani, Band, 2022), the initial promise is reported as finally delivered. According to Paes, quoted in the article,

the delay was caused by the discontinuation of his government when Crivella was elected Mayor of Rio in 2016.

It will be important in future research to analyse Eduardo Paes' government as Mayor of Rio from 2020 to 2024 from the perspective of the media coverage of the Olympic legacy. News published in 2022 on Rio's legacy points not only to the cyclical nature of the legacy discourse in the media, but also to Paes' reelection as an opportunity for the Mayor to rewrite the legacy narratives. As argued in section 2.2, legacy as a media event continues to evolve after the Olympic Games, not only because 'new legacies' are created but also because previous legacies are reinterpreted. This poses a particular challenge to the study of media events such as the Olympic legacy because determining a time frame for studying such phenomena becomes a complex task. For this thesis, the time limit was the date of conclusion of this study but this does not necessarily mean that the event had come to an end. In this sense, drawing on Maingueneau's statement that discourses are dialogic (2004, p. 52), the legacy discourse only exists in consonance with other discourses which, by reappropriating, expanding and reinterpreting the meaning of legacy, also expand the duration of the Olympic legacy as a media event.

5.9 The Government discourses: the analysis of official websites

To analyse official government publications on the legacy of the Rio Olympics, the website rededoesporte.gov.br and its archived documentation at http://arquivo.esporte.gov.br were collected, as discussed in Chapter 3, and fall within the Brazilian Government Corpus. Rio 2016 did not have its own webpage like London's queenelizabetholympicpark.co.uk. Both Brazilian websites were maintained by the Ministry of Sports until 2018. With President Jair Bolsonaro's election, in 2019 his government decided to abolish the Ministry of Sports, which was incorporated into the Ministry of Citizenship. Rede do Esporte, however, was still updated by the government's press office agency, producing articles on the Brazilian sport scene.

As the website is connected to the Federal Government, it also echoes the disconnection between Rio's Municipal Government and the Brazilian Federal Government. Eduardo Paes is cited in just one article entitled 'Cultural and mobility legacies will be hallmarks of 2016, say experts' (Rede do Esporte, 2015). The article reports on the book *Depois dos jogos: pensando o Rio para o pós 2016* (After the games: thinking post-2016 Rio) and cites the Mayor just at the end as a member and president of the C40, a global network of mayors created to find solutions to climate change. Other politicians, however, are mentioned in the articles, such as Leonardo Picciani, Paulo Márcio and Ricardo Leyser.

Leonardo Picciani was Minister of Sports from 2016 to 2018, appointed by the President Michel Temer. He is the son of Jorge Picciani, ex-president of the Rio de Janeiro Legislative Assembly (the state parliament), who was arrested twice for involvement in corruption schemes (Nitahara, Agência Brasil, 2021). In the articles, Leonardo Picciani is cited as the main representative of Rio's legacy, for example in 'Event at Barra Olympic Park celebrates one year of the Games and the legacy of the event to the country' (Rede do Esporte, 2017b). Picciani is quoted saying: 'the Games were our opportunity to reshape the city, to rescue Porto Maravilha and cover kilometres of BRT and underground'. He used 'our' when referring to the infrastructure constructed for the Games without mentioning Eduardo Paes. Although Paes was a member of President Michel Temer's Party (Dilma Rousseff's successor after the impeachment), he had publicly declared support for Rousseff (Globo, 2015), causing instability in his relationship with the Federal Government.

Another key political figure was Paulo Márcio Dias Mello. He was the President of the AGLO. He appears in most articles defending the legacy of Rio, as in 'Brazil shines in the opening of the stage of the world circuit held at the Olympic Park' (Rede do Esporte, 2017). The way he defended the legacy was mostly by showing how he had managed to make use of the facilities to host new events. In this article, Mello described the contracts signed with different federations to keep competitions taking place in the Olympic Park, like the tennis tournament cited in the text. Mello said to Globo in April 2019 that the AGLO had become essential, as Rio's City Hall was not able to maintain and repair the legacy infrastructure

(Gozzer, Globo, 2019). He argued that if he had not repaired the Velodrome roof – burned in the multiple fires that had hit the site – nothing would have been fixed. Nevertheless Mello as President of AGLO, despite his alleged importance, was fired from office in 2019, months before the planned closure of the company.

Ricardo Leyser, the fourth most mentioned person in the most frequently occurring bigrams in the Brazilian Government Corpus, was Secretary of High Performance Sport in the Ministry of Sports. Leyser appears on the website in connection with the Brazilian Medal Plan. According to the article 'Ricardo Leyser: The Games are not the point of arrival, but the point of departure' (Rede do Esporte, 2017c), R$ 1 billion had been invested in improving the infrastructure for the preparation of athletes. The project had paid a grant called *Bolsa Podio* to competitive athletes with high chances of winning a medal. Moreover, financial support had been invested in professional equipment and facilities around the country, not only in Rio. Leyser described at the end of the text what he called the 4 concepts of legacy that guided his work: wide, impacting all forms of sport; democratic, from the amateur to the professional athlete; national, reaching the entire country; and long lasting, going beyond 2016.

Bolsa Podio, however, is mentioned in only three articles in the Brazilian News Media Corpus, once in each outlet Globo, Folha and Estadão. Globo's 'At the end of Rio 2016, Government will keep investments in sports' (Globo, 2016) was published in August 2016. In the article, Leonardo Picciani is interviewed and he promised to keep the financial support for athletes as a public policy after the Games. According to the ex-Minister, 77% of Brazilian Olympic athletes were supported by the programme. *Bolsa Podio* is mentioned by Folha in the article 'Government cuts off 87% of the Ministry of Sports budget to 2018' (Folha, 2017) in a negative context. The programme was threatened by cuts in the government's budget. Estadao's article 'Rio's victims of violence try to reinvent themselves in canoeing' (Estadao, 2018c) tells the story of three people who became Paralympic athletes after having their bodies damaged by violence, such as gunshots. One of them, Luciano Meirelles, was considered an athlete with a high chance of getting a medal in Tokyo, due to his success in

recent years. Despite his performance, the article stresses that Luciano had to interrupt his training because his funding had been cut off.

The *Plano Brasil Medalhas* (Brazilian Medal Plan) was mentioned just once by Estadao in 'Former Sports Minister was fined for lack of legacy after the Rio 2016 Games' (Estadao, 2017a). According to the article, Ricardo Leyser and the former Minister George Hilton were fined for not presenting a Legacy Plan for the Games. The Federal Court of Accounts (TCU) considered the plan presented by the government as insufficient, as it did not specify costs in relation to maintenance and the responsible bodies to financially support these costs. In response to the allegations, Leyser and Hilton mentioned the *Plano Brasil Medalhas* as a concrete legacy of the Games.

Auto rendimento (high performance) is one of the most frequently occurring bigrams in the Brazilian Government Corpus (4th in ranking). This is a key point of difference with London, where elite athletes were not a priority in the Government legacy discourse. However, it points to a similar pattern encountered in London's analysis: the government emphasised programmes and projects that did not have the attention of the media, as in the International Inspiration Programme. *Auto rendimento* refers to *Bolsa Podio* as it was focused on high performance athletes. This pattern demonstrates a weakness in the government's communication strategies, which was not able to set the agenda – in the sense described by McCombs (2004, p. 2) – which means determining the most important topics to be discussed by the newspapers.[79]

Controle dopagem (doping control) is also among the most frequently occurring bigrams (12th in ranking) in the Brazilian Government Corpus. Most of the citations refer to the new Brazilian Anti-Doping Lab (LBCD) introduced for the Games. Again, the topic had little impact in the news about the Olympic legacy of the Rio Olympics. The only article to

---

[79] It should be noted, however, that municipal official websites were not included in the corpus due to their unavailability, which is a limitation in the analysis of the relationship between media and government discourses.

mention the laboratory was the already cited 'Environmental legacy: the great debt of Rio Olympics*'* (Soares, Globo, 2016) in which a series of legacy promises were checked by the journalists. In the article, first, the promise is mentioned: 'one of the main legacies of Rio 2016 candidature will be the creation of a National Anti-doping Agency'. The promise is followed by its 'status' in July 2016: 'the agency was created, but the Brazilian Anti-Doping Lab (LBCD) is temporarily unlicensed by the World Anti-doping Agency'.

Other bigrams such as *Arena Carioca* (Carioca Arena)*, Centro tenis* (Tennis Centre)*, Parque Barra* (Barra Park) refer to the facilities under the management of the AGLO, and were mentioned previously in the analysis of media discourse. They represent a shared point of interest between the government's official communication and the news coverage. This is expected not only because the Federal Government was responsible for the maintenance of these facilities, but also because they were among the most representative material-tangible legacies of the Rio 2016 Games. This is an example of a more successful alignment between government and media discourses, which did not occur much in the previous analysis of London 2012.

5.10 Activist blogs: where urban development becomes gentrification

The source used to study the activists' voices was the website rioonwatch.org.br. RioOnWatch was launched in 2010 (it was originally called Rio Olympics Neighbourhood Watch) by the organisation ComCat (Catalysing Communities). According to their website, the main purpose was to make the favelas' voices heard before the 2016 Olympics. Their news website is composed of articles produced by community journalists, local residents, international 'observers' and scholars interested in the role of favelas (slums) in urban development. Some of the articles published by the website are translations into Portuguese of articles published by international media outlets such as The Guardian or The Washington Post. The choice to produce a bilingual website – in English and Portuguese – highlights their interest in having a global impact by targeting an international audience. Activist blogs in relation to the hosting of mega-events like the Olympics are typically interested in people's

displacement caused by urban planning, which is a global phenomenon also evidenced in the British Activist Corpus.

A list of bigrams was produced from the text of 40 articles collected from the activist website and which constitute the Brazilian Activist Corpus. Unlike the news media and the government website, *copa mundo* (World Cup) appears among the most frequently occurring bigrams (12th in ranking). The reason for the event hosted in Brazil in 2014 being the 12th most mentioned is due to its similarity with the Olympic Games, regarding the processes of violent exclusion caused by the construction of the main venues.

The word cloud below shows the most frequently occurring words in articles published by activist websites. By looking at the image, words that have been rarely mentioned so far can be identified, such as *comunidade* (community), *moradores* (local residents), *vila* (village), *remoções* (removals), and *favelas* (slums).

Figure 39 – Word cloud produced with Voyant Tools using most frequently occurring words in articles published by activist websites

Studies have shown the negative urban impact of hosting these mega-events in Brazilian cities (Magalhaes, 2013; Castro & Novaes, 2014; Borges & Santos, 2018), mostly characterised by gentrification: increase in the value of properties; the removal of local residents; and the attraction of well-resourced individuals and businesses which leads to higher costs of living. The processes of removal of local residents and communities – like the Clays Lane example mentioned in the chapter covering London – happened before 2014 for the World Cup in several cities across the country and were repeated for the 2016 Olympics in Rio. One of the most high-profile cases of removal that was prominently covered by RioOnWatch was the case of Vila Autodromo.

5.10.1 Vila Autódromo: a glocal resistance

'The price of gold: one village's resistance sparked a global question: should the Games be abolished?' (Donahue, RioOnWatch, 2020) is the headline of an article published by the activist website in translation. The original was published by The Washington Post Magazine. The author, Bill Donahue, highlighted the fact that while most news stories had covered the impact of the postponement of Tokyo's Games on athletes' lives, the coverage had not given attention to other people:

> poor people who find themselves living in the wrong place, on land that the Games need for stadiums and parking lots. Poor people who are, per Olympic tradition, evicted and sent packing into a difficult new life (Donahue, RioOnWatch, 2020).

In Rio, Vila Autódromo was a community – or favela – located where the Olympic Park was to be constructed. It could have been just another removal caused by the Games, as historically documented in previous host cities, but the community's act of resistance caught the attention of the international press and became a symbol of the negative legacy of the Olympic Games. One of the movement's leaders, Maria da Penha, embodied the violence caused by the event: 'They smashed my face and demolished my home: all for the Rio Olympics' (Guardian, 2017). Maria da Penha was brutally attacked[80] by the police officers responsible for displacing the community, who broke her nose and left her face covered in blood. *Maria Penha* is the seventh most frequently occurring bigram in the Brazilian Activist Corpus (24 times). *Dona Penha* (Mrs. Penha) is also mentioned 12 times. RioOnWatch articles told her story. She moved to Vila Autódromo in 1994. When the City Hall offered her money to leave her house, her answer – 'not everybody has a price' – became the slogan of the resistance movement. Other bigrams related to the same story are *moradores Vila* (Vila residents)*, direitos humanos* (human rights)*, moradores favelas* (favelas residents)*,

---

[80] Ironically, Maria da Penha has the name of the Brazilian law created to protect women against violence and abuse. The law was named after a woman who became paraplegic after being a victim of her violent husband (Uchoa, BBC, 2016)

planejamento urbano (urban planning)*,* and *Barra da Tijuca* (Barra da Tijuca neighbourhood).

After the Vila Autódromo case, according to Donahue's article, the International Olympic Committee revised its questionnaire sent to the candidates for host cities, in which they now ask organisers to identify projects that involve removals and explain why these are necessary. The article mentions that in Tokyo, 'only' around 300 people were removed from their houses for the Games.

Contrasting the activists' blogs with the news media, in the latter, the connection between Vila Autódromo's struggle and the words 'Olympic legacy' was rare. Only five articles in the news corpus mentioned Maria da Penha or Vila Autódromo - three published by Folha and two by Globo. In 'Olympic village and the Vila Autódromo*'* (Boulos, Folha, 2016), the opinion article discussing the removal of families from the Olympic Park territory concludes by saying that the Olympic legacy for the poorer was evictions, real estate speculation and Rio's militarisation. In '*Cidades e Soluções* follows works in Rio for the 2016 Olympics*'* (Globo, 2013), a review of the main projects for the Olympics is presented. Under 'Olympic Park', the article mentions the eviction of Vila Autódromo residents as a conflict. Following the residents' protests against the violent process led by the Mayor, Eduardo Paes is quoted saying that this was 'a solved issue'. Vila Autódromo is also not the main topic in 'One year after Rio 2016, Olympic arenas are underused*'* (Globo, 2016b). In the article, the fight for land is mentioned only *en passant*.

Based on this analysis, it is clear that Vila Autódromo was largely excluded by the media in the Rio Olympics legacy narrative. While it was the main topic highlighted by activists, and also had a global impact, local Brazilian media did not cover the topic as a negative consequence of the 'Olympic legacy'.[81] While people's displacement might not be associated with legacy in the news articles, the activists' blogs explicitly make this association. The tactics of making the struggle of the community of Vila Autódromo be encapsulated within

---

[81]It is, however, important to acknowledge that this story may have been covered in articles not explicitly focused on the Olympic legacy given the limitations of the data collection process.

the legacy discourse is a prime example of the construction of counter-discourses. Counter-discourses are intrinsic to the the concept of 'discursive events' (Hepp and Couldry, 2010, p. 11), where the hegemonic discourse is under dispute. If legacy is a central part of the IOC mission and is associated – by the IOC – with a positive impact caused by the event, activists construct the argument that gentrification is also a result of the event, and therefore, a legacy. As stated by Maingueneau (2004, p. 52), discourse is action, and activists are particularly committed to making use of the powerful nature of language to challenge hegemonic discourses.

5.11 Contrasting Brazilian media outlets

As in the analysis of the British coverage of London 2012's legacy, this section will look at the differences in the coverage between the different news outlets. A brief introduction to the topic regarding the Brazilian news media context was outlined in section 5.1, where the economic concentration of news outlets in Brazil was discussed. This section will present an overview of the companies and their characteristics, followed by a comparative analysis to understand how their narratives differ.

Globo Group is the biggest media company in Brazil (BBC, 2023). The company was founded by Irineu Marinho in 1925. During the Brazilian military dictatorship (1964-1985) his son, Roberto Marinho, founded TV Globo, in 1965, which became the largest Brazilian Broadcast company. Today, the group owns TV and radio channels, streaming platforms, newspapers, magazines, news websites, publishers and film production companies. Among the 16 most accessed offline news media in Brazil, according to the Reuters Institute Digital News Report (2022), four are part of Globo Group: TV Globo (1st), O Globo (4th), Globo News (7th) and Jornal Extra (15th). Its monopoly is also reproduced online, with Globo News/G1 being the most accessed news website in the country and O Globo Online the fourth. Globo Group remains owned by the Marinho family. The concentration of media ownership is highlighted in the Reuters report, with Globo Group being one of the main

players. TV Globo was a Rights Holder with permission to broadcast the 2016 Rio Olympic Games.

Folha Group is a media conglomerate having as its main product the newspaper Folha de São Paulo, which started to circulate in 1921. In 1962, it was bought by Octavio Frias de Oliveira, whose son, Luiz Frias, was still the owner of the company in 2022. Folha Group owns newspapers, news websites, publishers, a statistics and survey institute and its most profitable business (Globo, 2020) PagSeguro, a digital payments and credit company. The newspaper – under paywall – is the 11th most accessed offline news outlet according to the Reuters report [82] (Digital News Report 2020, p. 90). Online, its product UOL is the second and Folha online is the eighth.

Estadão – or O Estado de São Paulo – was founded in 1875 during the colonial period as A Província de São Paulo, changing its name after the Republic transition in 1889. The newspaper is part of Estado Group, which also owns a news agency, radio and a record company. The group was founded and is still controlled by the Mesquita family. It was the 16th most accessed news website in Brazil according to the Reuters Report.

5.11.1 Content analysis

Folha presented the most critical news coverage of Rio's Olympic legacy. The majority of the news articles collected for this research reported on negative aspects of the legacy. Among Folha's main topics were funding problems, the government's financial struggle and inefficiency in delivering the legacy. Folha was mainly focused on the political reactions to the legacy challenges. In the word counts for Folha, the word *milhoes* (million) appears as the fifth most frequently occurring, mainly in the context of money as in 'With debt of R$ 130 million, Rio 2016 asks governments for help*'* (Folha, 2017b). The words *governo* (government), *prefeitura* (city hall) and *ministerio* (ministry) also appear among the most

---

[82] I refer to this report (Digital News Report 2020) in Chapter 3.2, as a document utilised to define the data collection and corpus creation for this study based on news media outlets popularity on the internet.

frequently occurring. *Paes* - from Eduardo Paes - is 17th in the ranking, being mentioned 73 times.

Globo, however, presented a more diverse coverage, focusing not only on the problems but also echoing Eduardo Paes' expectations of an urban transformation as a result of the Games, for example in 'School named after Socrates will be part of the Olympic legacy project' (Globo, 2012c). While Folha's coverage of Rio's legacy was mainly published after the Games, Globo dedicated more space to the topic before the event.[83] Among Globo's most frequently occurring words is *atletas* (athletes). The word was not among the 50 most frequently occurring bigrams in Folha. Athletes are mentioned by Globo mainly in the context of being a product of Rio's legacy, in terms of investment in their careers, or the emergence of new figures to be admired by the public, as in 'Olympic legacy: emergence of Idols is an incentive to practice sports' (Barros, Globo, 2016). Facilities such as *Arena* and *Velodromo* are also among the most frequently occurring unigrams. *Milhoes,* however, is only the 21st most frequently occurring. It is still relevant (mentioned 209 times) but proportionally less than in Folha.

Estadao also presented more diversity in its coverage than Folha. Among the 15 articles published from 2009 to 2015, only one presents a critical and negative framing of Rio 2016: 'Rio suffers from "violence, bribery, favelas and decrepit Maracanã", says British newspaper' (Estadao, 2012b). The article cites a publication from The Times, highlighting the negative perspective of the international media in contrast with the Brazilian optimism for the Games. From 2016 onwards, Estadao's coverage presented a balance between the positive aspects of the games - sometimes based on promises and less concrete deliveries, as in 'Barra Olympic Park will have a national sports museum' (Estadao, 2018), and at other times based on

---

[83] Globo's first article in the corpus was published in 2010. <https://extra.globo.com/noticias/extra-extra/relator-do-plano-diretor-inclui-legado-olimpico-372052.html>. Archived at <https://web.archive.org/web/20230612145750/https://extra.globo.com/noticias/extra-extra/relator-do-plano-diretor-inclui-legado-olimpico-372052.html>.

materialised legacies, as in 'Finally, Olympic legacy for Brazilian sport arrives in Deodoro' (Estadao, 2018b).

A more detailed comparison in terms of sentiment will be conducted in the next chapter. What this qualitative close reading analysis, in combination with the quantitative approach, has revealed is that Globo, the only Rio de Janeiro based outlet and official broadcaster of the Games, and Estadao dedicated significantly more space to the coverage of the legacy in comparison to Folha, since Rio was elected as host city. Both of them reproduced - or produced - the discourse of positive expectations. Folha, however, was mainly focused on the negative impacts of hosting the mega-event.

A characteristic that the narratives of legacy in the different media outlets have in common is the lack of a published legacy plan to guide the journalistic work of watching the government. This absence transformed the news coverage into a set of oral promises made by politicians, primarily focused on the infrastructure constructed for the Games.

5.12 The g*ambiarra* Games

The Rio 2016 opening ceremony started with an impressive scene of waves – resembling the beach – formed by pieces of glossy paper that changed their form as people folded and unfolded them. The concept behind the scene was described as *gambiarra*. The word, referred to as a MacGyver approach by Reuters (Milliken and Haynes, 2016), means improvising solutions, or what the ceremony narrators described as 'the art of making something right out of almost nothing'. *Gambiarra* was not just a theme to shed light on the lack of resources available to produce the ceremony; it was a metaphor for a Brazilian way of life, profoundly marked by the need to overcome precarity by being creative. While the ceremony placed the concept in a beautiful imagery of resilience, *gambiarra* extends beyond its first meaning.

The Priberam Dictionary of Portuguese Language (2022) describes *gambiarra* in five ways. Two of them are important here: '1. Electric extension, with a long cable and a lamp at its

end, which allows light to be brought to remote places' and '5. Improvised solution to solve a problem or a need'. While the second (number five) is closer to the meaning of the concept in the context used in the opening ceremony, the first helps to understand its origin. *Gambiarra* is a solution for bringing energy further by connecting one cable to the other. This is obviously a dangerous solution as it can cause accidents. Based on this origin of the word, it is possible to say that *gambiarra* is not necessarily an inspiring talent to overcome difficulties but a dangerous, poorly made, inconsistent and provisional solution for problems of daily life. This less beautiful meaning is a useful metaphor to understand the narratives of the Olympic legacy in the media coverage of the Rio Games.

The article 'Rio Justice determines presentation of Olympic legacy plan by August 4' (Yahoo Brasil, 2016) was published on 20 July 2016. It reports on a decision made by the Brazilian Court that the Rio 2016 Games authorities had to present a legacy plan with a deadline set for the eve of the Opening Ceremony. According to the judges' decision, the government should present a document detailing all the facilities constructed for the Games, the investment made and who was going to take responsibility for their maintenance. The article says that 'Rio Mayor Eduardo Paes insists that the city's legacy plan was presented about a year ago, but so far it is unclear who will assume the financial responsibilities for maintaining Olympic facilities after the Games, nor what sporting projects will be carried out in permanent arenas'. The text continues with Eduardo Paes' description of the 'legacy plan':

> The handball arena will become four schools, the aquatic stadium will become two gymnasiums, the Carioca 3 arena will become a school, the Carioca 2 will become the Olympic training centre and the Carioca 1 "we will grant it to the private sector". "The velodrome and tennis arena will have a social function and stage for events and Deodoro has a social role," he added. (Yahoo Brasil, 2016)

It was only in March 2022, six years later, that the handball arena (Arena do Futuro) began to be dismantled for the construction of the promised schools (Diario do Rio, 2022). This case illustrates what was repeated over and over again in the news articles: the Mayor's

declarations were taken as 'the legacy plan' and journalists stuck to them to narrate the future – what is going to be done – and the past – what was promised.

At the beginning of this chapter, an article reporting on Nick Clegg and Jeremy Hunt's visit to Brazil was cited. Their advice, based on what they had learned from London's experience, was that Rio should start planning their budget in 2011, as early as possible. The lack of planning became in the following years the government's Achilles' heel and was widely discussed by the newspapers.

Compared with London's analysis, Rio's news articles reproduce a sense of *gambiarra*, in the sense of an improvised legacy. But my argument here is not only regarding the coverage of the failed legacy represented by the abandoned venues (the white elephants), the accidents in poorly constructed infrastructures, the green pools and the dirty Guanabara Bay. These were obvious results of a *gambiarra* way of managing public policies. My argument goes beyond that and focuses on the narratives of the news coverage. As the media did not have a document on which to base their coverage of the legacy, Rio's legacy plan was constructed through and with the media. I say 'through the media' because they gave voice to the Mayor's promises and reported on what he said would become the legacy in the future. But it was also 'with the media' because it became a 'place of reference' (Vizeu, 2009) for the legacy, precisely due to the lack of official documentation. The Rio 2016 official website was taken down. The Ministry of Sports was dismantled by President Bolsonaro and became part of the Ministry of Citizenship. The disconnection between Rio's Municipal Government and the Federal Government further impacted the quality of information produced by the website Rede do Esporte. Rio does not have a website specifically for the Olympic legacy. As a result, the news media websites are the main sources for accessing, studying and understanding the legacy of Rio and its history.

The Rio Mayor's oral promises, as reported in news articles, however, are lacking in the narrative order that can be found in official documentation. In the case of London 2012, it was possible to easily compare the delivery of the legacies with the legacy plan that was published before the Olympics. For Rio, it becomes more complex as what is or is not a

legacy was defined by the Mayor and subject to change. This brings a disorganised characteristic to Rio's legacy narrative. It is hard to identify when a project plan emerged, when it was mentioned for the first time and consequently to track its development. Sometimes, projects were cited as delivered before being announced as legacy plans. This is the case of the Socrates School, mentioned earlier in the analysis, that was delivered in 2012 and announced as part of Rio's legacy by the Mayor Eduardo Paes later. Paes exploits the blurry temporal and semantic boundaries of legacy as a media event, as discussed in Chapter 2, to continue to introduce new 'legacies' and to expand the media event for his own political gains.

The media coverage in general was mainly focused on the material legacy – more specifically the venues constructed for the Games. In contrast, the government's plan of investing in high performing athletes had little influence on the news agenda, nor did the activists' attempts to highlight Vila Autódromo's struggle.

The extremely positive article published by Globo in 2016 'Olympic heritage for Cariocas' (Intrator, Globo, 2016) replaces the word legacy with one that fits more aptly with the way the Olympic legacy of Rio was developed and narrated. The article says that there is no doubt that the facilities constructed will be a true legacy for Rio. In the title, however, the word used is 'heritage'. As mentioned in Chapter 2, heritage carries a stronger sense of what was left behind after the Games (MacAloon, 2008). Rio's legacy narrative in the media is generally more aligned with this understanding of legacy that focuses more on the past than the future. This is evidenced by the publication of Rio's legacy plan document in June 2017 after the Games' conclusion, highlighting most clearly the absence of future legacy planning in the preparations for the Games.

5.13 Conclusion

This chapter combined the quantitative and qualitative analysis of news articles published by Brazilian media outlets on the legacy of the Rio 2016 Olympic Games – as well as the two

corpora composed of government and activist articles –, while also acknowledging the methodological challenges of finding and working with Brazilian and Portuguese-language sources.

The analysis highlighted the positive discourse of Rio's authorities as presented in the media regarding their expectations of transforming the city through the Games, focusing in particular on the material – tangible – legacy and urban infrastructure. The media also focused on the challenges faced by Rio in order to prepare for hosting the Olympics, primarily with regards to the decontamination of Guanabara Bay – with special focus on the discourse of promoting sustainable Games–, the fires in the Velodrome, and the building and privatisation of infrastructure. The fires shed light on the precarity of the material legacy, while privatisation dialogues with a widespread discourse that private management of legacy would avoid corruption schemes.

As in the case of Boris Johnson, the political trajectory of Eduardo Paes, Rio's mayor, was salient in the media coverage and he had a dominant influence over the legacy discourse. This finding shows how politicians, especially heads of host cities, shape legacy narratives, which could lead to political career benefits, such as being recognised as responsible for promoting urban development and other transformations in the host cities.

The analysis of official government communications highlighted how the Brazilian and Rio governments disputed the legacy narratives, which shows the interest of the Brazilian presidency and being recognised as responsible for delivering the legacy as well. While President Lula da Silva was involved in the bid and celebrated the election of Rio as host city of the 2016 Games, he and the following Brazilian presidents lost visibility in the media narratives of legacy in favour of the Mayor.

The analysis of the government websites also showed the failure of the Brazilian government to influence the media agenda in relation to its initiatives to support Brazilian athletes. While an investment of public money was being made by the government on this particular programme, little attention of the journalistic coverage was given to the topic of elite athletes.

To conclude, this chapter presented an analysis of the response of activist blogs to gentrification and their interaction with global discussions on the impact of hosting the Olympics, with special focus on the case of Vila Autódromo. Results showed that the struggle of communities in Rio was not an isolated phenomenon, but rather interconnected with global discussions on the transformation of urban centres worldwide. However, while activists emphasised the negative impact of the Games in local communities as part of a – negative – legacy, the media did not associate people's displacement with the Olympic legacy of Rio 2016.

In this chapter, I also analysed the economy of the Brazilian media outlets and how their content differs from one to the other. This is of particular importance as the analysis highlights the significant concentration of media ownership in Brazil and its dependency on private capital. Understanding how this concentration translates into content production would require further studies that consider documentation beyond newspaper articles, such as information about companies that pay for advertisement in those media channels. Still, the analysis of the media landscape helps understand the massive coverage produced by Globo. The outlet was, among the analysed sources, the only company placed in Rio and the official accredited broadcast of the Rio Games in Brazil.

Finally, I suggested the concept of a *gambiarra* Games, as a way to understand how the legacy of the Rio 2016 Olympic Games has been narrated. This concept of 'gambiarra' Games highlights the value of studying the news media narratives of the Olympic legacy. While scholars such as Gratton and Preuss (2008), Poynter and MacRury (2012), Cashman (2005), MacAloon (2008), and Girginov (2018) have contributed to the scholarly debate on the Olympic legacy by analysing a diverse range of documents, including official documentation produced by the IOC and local organisers, and therefore contributed to understandings of how legacy is constructed, none of the authors have dedicated significant attention to news media. The discursive mechanisms of news articles, which are evidenced here through the concept of *gambiarra* Games, also produce legacy. Particularly in a scenario where there is a lack of official documentation on legacy planning, as was the case of Rio

2016, studying the news media narratives allows us to reconstruct the legacy of such mega-events. It is a challenge to look back at the Rio 2016 Olympics Games and list what is or is not a legacy of that event. This study, therefore, adds to the scholarly debate on the Olympic legacy, and the legacy of mega-events more broadly, by providing further insight into how discourses of legacy circulate, how they are produced, as well as how they can be reconstructed from limited and partial primary sources. The analysis highlighted in particular the importance of particular stakeholders, such as the Mayor of Rio, as central figures that through their engagement with the media hold the power to dictate what legacy is.

In the following chapter, a deep analysis of the sentiments discursively expressed in the texts published by the Brazilian and British media will be carried out. The use of SA aims to enrich the qualitative close reading presented in this chapter by looking at the legacy narratives from the perspective of sentiment and affect. While some preliminary conclusions regarding the more positive coverage by Globo and Estadao were presented in this chapter, it is necessary to go further to analyse in more depth the role of positive, neutral and negative sentiments in the discourses.

Chapter 6: Language, sentiment and legacy: an affective dispute

In this chapter, Sentiment Analysis classifiers are used to explore the dataset of news articles covering the Olympic legacies. Sentiment outputs are interpreted and analysed in relation to the discourse analysis in order to enrich our understanding of how media outlets covered the legacy of the Games.

6.1 Why does sentiment in the news matter?

As introduced in Chapter 1, De Hollanda and Medeiros (2020) analysed 84 news articles on the coverage of the Olympic Games in Rio through close reading of the texts. The authors first divided the corpus into three segments of time – before, during and after the Games – and manually classified each of the articles based on whether their content was positive, negative or neutral. Their paper concluded that the Brazilian media coverage of Rio 2016 was more negative before the event and became more positive afterwards. The results presented in this chapter, however, have found the opposite to be true.

It is important to highlight that the articles utilised in De Hollanda and Medeiros' study were not specifically covering the 'legacy' of the Games, but rather covering the event 'Rio 2016' in general. When considering this broader scenario, it is expected that news articles on competitions, medals, victories (and losses), and festivities among other topics influence their conclusions. In the case of this thesis, however, the specific framing of legacy tends to reduce the interference of these other adjacent aspects of the event. I use 'tends' because in some articles it is possible that, despite mentioning legacy, aspects of the text that contribute to the sentiment score may be focused on something else that is not the legacy per se.

De Hollanda and Medeiros' work shows researchers' interest in sentiment in the news. The particular contribution that both the authors and I seek to make when looking at sentiment is to understand the tone in which news stories are narrated. Again, as stated throughout this

thesis, if the media is understood here as a 'place of reference', the question that emerges is: how does negative or positive coverage of such an event impact the way this event is understood and interpreted by the audience? This thesis cannot answer this specific question because it does not look at audience reception. But what this thesis can do is to provide a comprehensive analysis of how sentiment is discursively expressed in the media.

De Hollanda and Medeiros' took the approach of close reading 84 articles, analysing and categorising them manually. This thesis instead considers a much larger corpus of 1271 articles, which gives a broader insight into the media coverage of the Olympics in comparison to a smaller corpus. By using algorithms to classify sentiment in texts, my aim was to be able to explore and understand the broader dynamics at play, while still having some control over the method and its outputs. This is why I did not apply sentiment to the entire dataset, but only to the sub-curated one, which is composed of articles whose main topics of discussion was annotated (see Chapter 3.4) . Future work would benefit from exploring ways to adapt this pipeline to analyse sentiment in the media coverage of other events and on an even larger scale.

To analyse the sentiment of the articles, I follow a similar approach to that taken by De Hollanda and Medeiros (2020), as well as Haw (2022), by dividing the data into before, during and after the Olympics. This temporal analysis contributes to the understanding of how sentiment is transformed over the years and how it is impacted by the staging of the event.

In the following sections I present the results of the experiments with SA and the interpretation of what sentiment in news articles can tell us about the concept of legacy.

6.2 Measuring sentiment

By plotting the sentiment labels in charts, it is possible to have an overview of the results. Figures 40 and 41 illustrate the percentage of positive and negative labels assigned to news

titles and body texts. The name of the outlets is followed by letters that indicate the event they cover: 'L' for London and 'R' for Rio.

Figure 40 – Comparison between the percentage of POSITIVE labels assigned to news titles versus body text
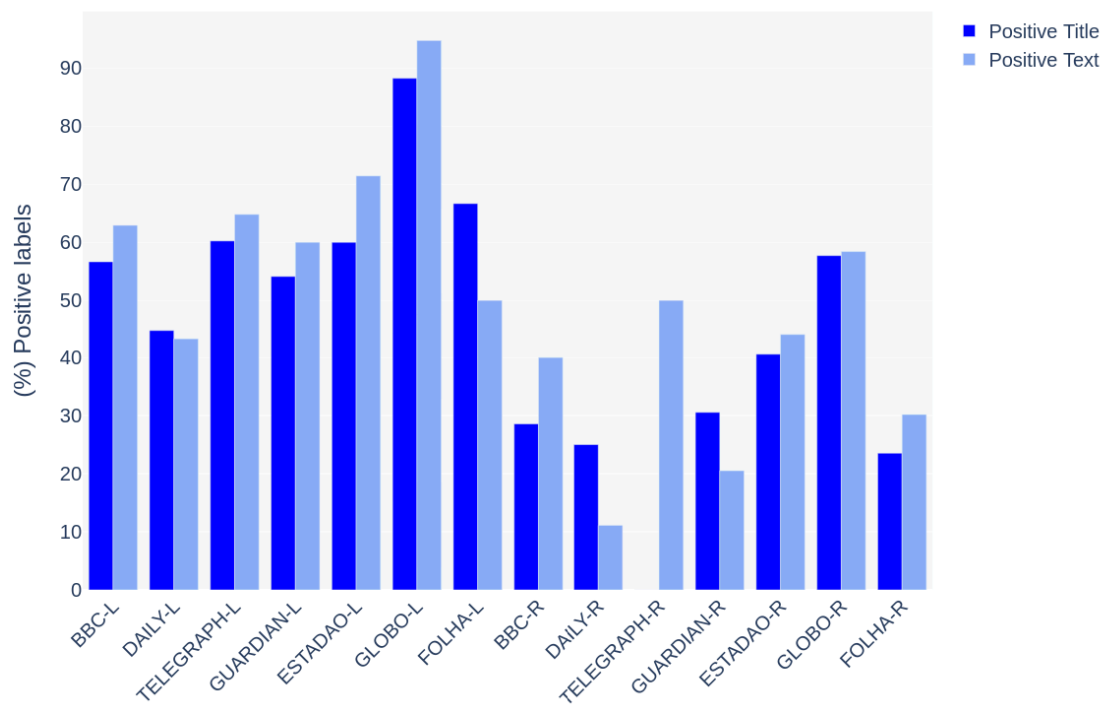
Figure 41 – Comparison between the percentage of NEGATIVE labels assigned to news titles versus body text
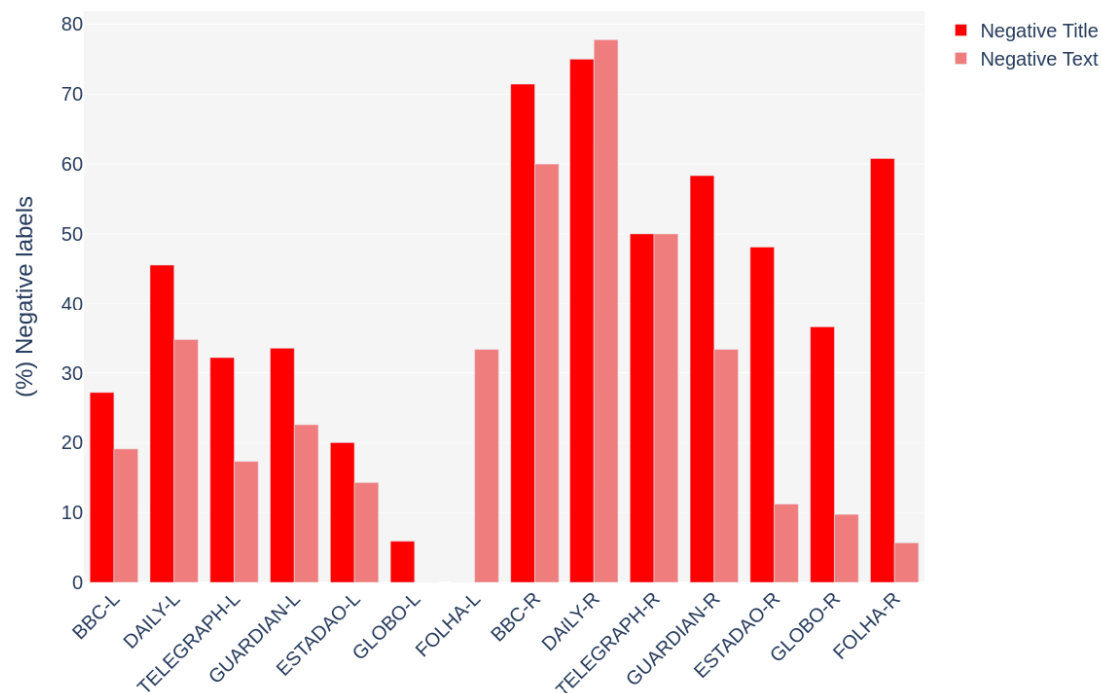
Charts in Figures 40 and 41 show that news titles tend to be less positive and more negative than their respective articles. Considering the particularities of the text type used as data here, titles have fundamental importance. What today is often referred to as 'clickbait' (Scacco & Muddiman, 2016) in online news is a result of an old strategy widely spread in the newsrooms. Rayner et al. (2001, p.227) describe news headlines as 'hooks' aiming to grab the eyes of the audience. For the authors, however, rather than just call the attention to the article, news titles must maintain a certain level of mystery, creating a problem to be solved in the body text. Negative headlines, such as 'Rio Olympics could be under threat as Brazil riots continue over political unrest (Parry, Daily Mail, 2016)', work particularly well to both catch the attention of the reader due to the use of the word 'threat' – how serious is the threat? – as well as for raising uncertainty and, therefore, creating a certain level of mystery: will the event be affected by the riots or even cancelled? Although not recent, this logic of convincing the reader to buy the story has become central to the production of online news whose audience is calculated based on the number of visitors and clicks received by the

website. Many online newsrooms assign to a 'homepage editor' the function of producing creative headlines to make the content more attractive.

Corroborating this understanding, Figure 42 shows how positive and negative labels are assigned significantly more to news headlines than neutral ones. Figure 43 illustrates the sentiments of the body texts, where the number of neutral labels increases, especially for the last four outlets: The Guardian-R, Estadao-R, Globo-R and Folha-R. Although titles for these outlets were more negative, the sentiment of the articles' body text does not correspond to the titles.

Figure 42 – Comparison between the percentage of positive, negative and neutral labels assigned to news titles
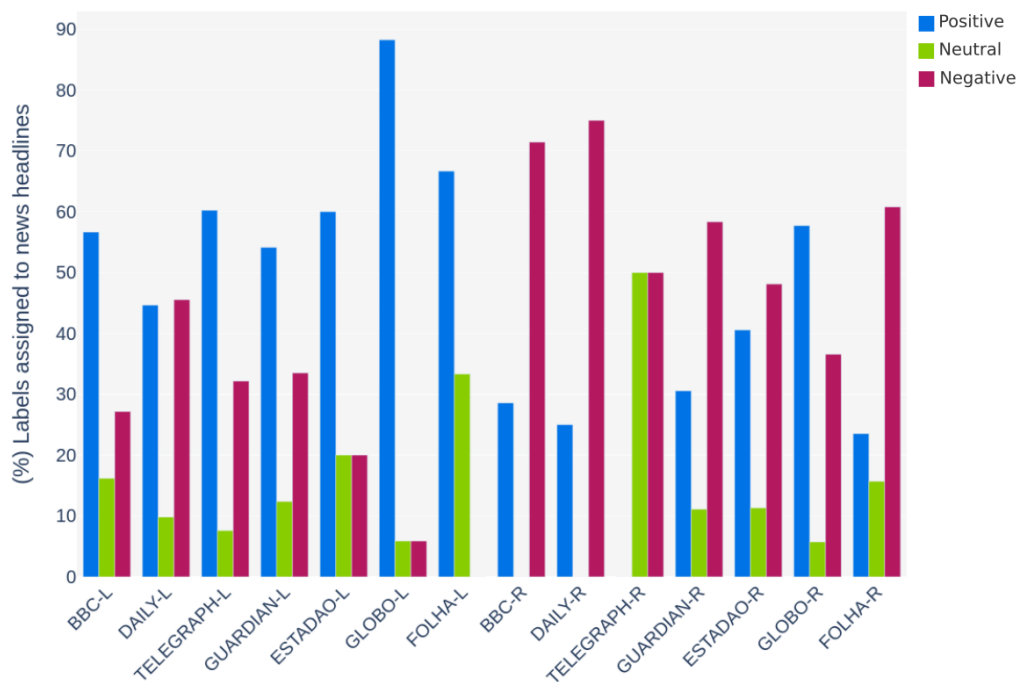
Figure 43 – Comparison between the percentage of positive, negative and neutral labels assigned to news articles' body text
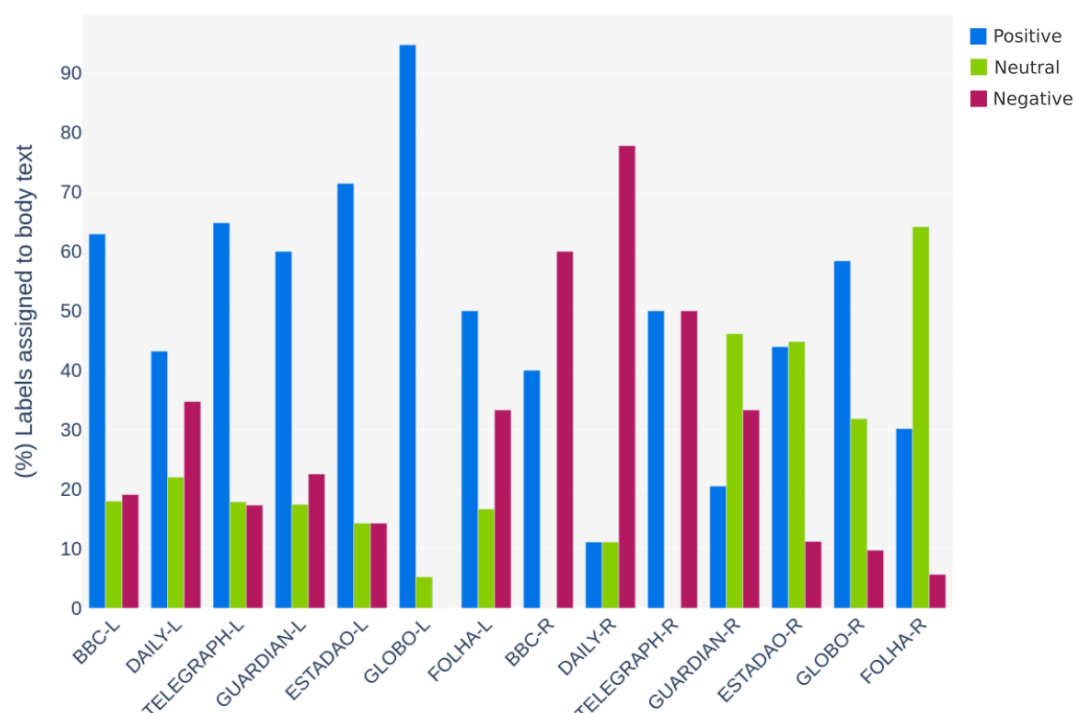
By comparing news headlines' sentiments, the contrast between the two events is evident: the media coverage of London's legacy was significantly more positive, while Rio's was more negative. Regarding London, the British media was more 'critical' than the Brazilian media. Estadao, Globo and Folha presented a very low number of negative headlines when referring to the London 2012 Olympics. However, when covering Rio, the British media had a higher level of negative news titles that was comparable with or, sometimes, even more negative than the local Brazilian outlets.

There is a potential bias of news article headline sentiment towards the negative. Rozado, Hughes & Halberstadt (2022) applied sentiment and emotion analysis to 23 million news headlines from 47 news media outlets in the United States and came to the conclusion that there was an increase in negativity in headlines across written news media from the year 2000 until 2019. Although a similar study was not found covering the British and Brazilian media, both countries were mentioned in the 2022 Digital News Report published by the Reuters

Institute for the Study of Journalism (Newman, 2022) as places where news avoidance had doubled over the last five years. Among the main reasons for residents of these countries to 'avoid' reading the news is that news had a 'negative effect on their mood'. There is no evidence within the report to confirm that people's mood had been affected because news was becoming more negative, as many other factors may have had an impact on this phenomenon such as the recent political crises in both countries.

6.3 Interpreting sentiments

The analysis of these sentiment outputs, combined with a qualitative close reading approach, confirms the hypothesis that Olympic legacy narratives are embedded in a utopian-dystopian dichotomy. As stated by Girginov (2018), the concept of Olympic legacy is a social construction that promotes 'specific visions of what the desired development should look like' (Girginov, 2018, p. 196). It is an idealised form of legacy that I refer to when using the term utopian narrative. This phenomenon is particularly promoted through official government communication strategies that emphasise the catalyst potential of the event for urban development. On the other hand, the fear of not achieving the promises, evidenced by the use of words such as 'white elephants', narrates the chaotic effect of the Games. This dystopian idea became prominent in expressions used by authors like Zimbalist (2017), who initiated his book's introduction about the Rio Olympics with the sentence 'welcome to hell'.

The analysis highlights this largely 'utopian' view of the Brazilian media in relation to London 2012, especially by Globo, while the British media primarily focused on the negative aspects of Rio's legacy, largely silencing other perspectives. The following table shows some examples of news headlines published by the British media covering the legacy of the Rio Olympics and the Brazilian media covering the legacy of the London Olympics.

Table 8 – List of news headlines published by the British media about the legacy of Rio and news headlines published by the Brazilian media about the legacy of London

News headlines from the **British media** about the legacy of **Rio 2016 Games**	News headlines from the **Brazilian media** about the legacy of **London 2012** Games
Rio's Olympic legacy a 'huge disappointment' (Davies, BBC, 2017)	*London 2012 will be the model for Olympic future, says Rogge* (Globo, 2012)
These haunting photos of abandoned Olympic sites tell their own stories (Varley, BBC, 2018)	*Six years later, the wonderful legacy of London 2012* (Pussieldi, Globo, 2018)
Rio 2016's venues abandoned and vandalised (Wade, Daily Mail, 2017)	*A blessed legacy* (Sousa, Globo, 2013)
Rio Olympics could be under threat as Brazil riots continue over political unrest (Parry, Daily Mail, 2016)	*London admits struggle to get Olympic legacy* (Estadao, 2011)
Zika crisis and economic woes bring gloom to Brazil's Olympic buildup (Watts, The Guardian, 2016)	*Final cost of the Games must burst* (Folha, 2012)
Rio's Olympic legacy lies in ruins - even if the IOC continues to insist otherwise (Brown, Telegraph, 2017)	*Olympic Park is returned to London and will receive almost BRL 1 billion renovation* (Folha, 2012b)

In the table above, there are two critical headlines published by the Brazilian media on London's legacy: 'London admits struggle to get Olympic legacy' (Estadao, 2011) and 'Final cost of the Games must burst' (Folha, 2012). Both headlines related to the high costs of hosting the Games and were published by Estadao and Folha. The interest of the Brazilian media in the cost of London 2012 is particularly interesting as its coverage functions as a preparation for Rio 2016. Watching London's performance and 'struggle' can be read as an alert about how these news outlets would cover the future Rio event and what the most important concerns would be.

6.4 The scepticism of Londoners and the optimism of Brazilians

Plotting the results in a chart divided into three parts representing the sentiments before, during and after each event facilitated the analysis of how positive, neutral and negative results varied overtime.

Figure 44 – (%) Variation of sentiment for news headlines before, during and after each event by outlet. Timeframe for London 2012 Olympics: before (2004 to 2011), during (2012), after (2012 to 2020). Timeframe for Rio 2016 Olympics: before (2009 to 2015), during (2016) and after (2017 to 2020)
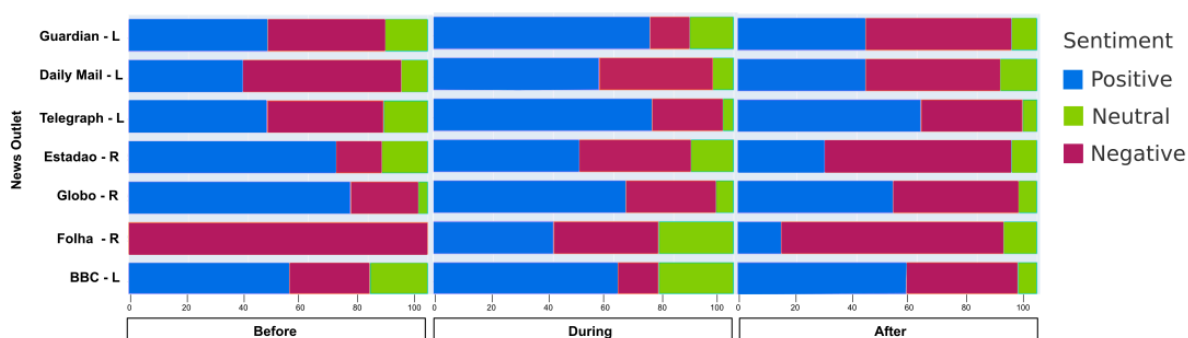
There are some important characteristics of each of the three time periods mentioned above that must be taken into account. 'Before' contains mostly the expectations for the Olympics. News covering the planning and delivery of the first sports facilities were in focus. However, besides shedding light on the preparation, news also predicted possible effects of the Olympics as in 'Olympic "legacy" expected in Kent*'* (BBC, 2005). 'During' is composed of news covering the delivery of the first facilities as in 'Olympic legacy: Rio community receives volleyball training centre*'* (Globo, 2016), as well as issues faced by the local organising committee to finish preparations on time. 'After*'* contains those articles that review the promises and compare them with the present reality as soon as the event finishes, as in 'After five months, Olympic legacy presents more negative points than positive*'* (Estadao, 2017).

By looking at the British media coverage of London in Figure 44, it is evident that there was an increase in positivity during the event in comparison to before it. The euphoria promoted

by the Games impacted significantly news that was less critical while the event occurred. After the event, however, the number of positive news titles dropped and negative stories increased.

For the Brazilian media there is a pattern followed by Globo and Estadao. Both presented a higher number of positive headlines before Rio 2016 that then reduced over time while the negative ones increased significantly. Folha, however, presented an anomaly, being only negative before, slightly more positive during and very negative again after the event.

The variation of sentiment over time points to an anaesthetising effect in the press during the Games. The agenda is impacted by the volume of news about sports, with the legacy being covered more from the perspective of what has been delivered and recently inaugurated; or even hopes and promises for the near future.Comparing the behaviour of the media with public opinion, it is possible to see a similar pattern. A survey conducted by Sesc RJ/FGV Projetos two months before the Rio Olympics showed that more than 60% of Rio citizens believed in the success of the event (Galdo, 2016). Another survey run a year later in 2017 by Datafolha (Folha, 2020), however, revealed that 70% of Rio citizens evaluated the Olympic legacy as negative.

Regarding the 2012 Olympics, although two-thirds of Londoners said that they expected to pay for the additional costs of hosting the Games, 69% supported the event (BBC, 2006). The number changes significantly, however, when the broader national scenario is observed. Around 64% of the British public responded to a poll saying that although they see the positive impacts of the event for London, they do not see the same for themselves (Gardiner, 2012). Dolan et al. (2016) compared the Londoners' sentiments after the Games with Parisians and Berliners and concluded that London citizens were significantly happier during the Games than the others, but that this sentiment returned to normal levels in the year after. As pointed out by Hiller & Wanner (2018), the measurement of public opinion about the Games is impacted by a diversity of sentiments that coexist and transform over time, which makes its study challenging.

6.5 Legacy and desire

In section 2.4, I presented arguments to sustain the idea that legacy and desire are intrinsically related, particularly in relation to the city. Urban development was central to the legacy discourse as demonstrated in the analysis presented in Chapters 4 and 5. It is evidenced in the British media coverage of London by the newspapers' obsession with the Olympic Stadium. In the Brazilian coverage of Rio, it is even more relevant, as the Rio Mayor's main stated interest was to use the event as a development catalyst. From the beginning, Paes argued that Rio was the best host city for 2016 not because it was ready, but on the contrary, because it was the one in greatest need of investment, in comparison to the other candidates. The construction of sports facilities and schools, as well as the intention to clean Guanabara Bay, were among the main topics discussed in the analysed articles.

The concept of legacy in this sense operates as what David Harvey called the 'object of utopian desire' (Harvey, 2012, p.16). The desire for a better city, with more green areas, more social spaces, more sporting and leisure facilities, better schools, more affordable homes, etc. is fed by the legacy promises of making use of unprecedented resources to transform the host cities in a very short period of time. With the city being a place of desire per se – that is a place for constructing new forms of living and existing in the world – what these promises do is engage local residents in the utopian vision.

Utopia, however, is not taken here with negative connotations, as an illusion. Utopia, according to the Uruguayan writer Eduardo Galeano (1994), is what keeps us moving forward. Utopian desires are constantly moving away from us the closer we get to them. Therefore they have the important role of keeping the flame burning. There is no radical change without desire. As pointed out by Girgnov (2018), legacy follows the motivational framing as stated by Snow and Benfored (1988), which in social movement theory is what produces action to overcome fear. That is how protestors decide to place their own bodies in vulnerable positions to defend their causes. This motivational framing is also what makes

legacy a key element of Olympism and, consequently, works as a legitimating narrative for the high costs involved.

The other dimension in which desire is manifested in the context of legacy is through the Olympic Movement's aim to inspire, motivate and guide the actions of entities committed to make the games. Here, I focus on discourses that aim to produce desire. In the analysed articles, this is evidenced in the focus of the British media on the promises of engaging young people in sporting activities. As shown in Chapter 4, there was a widespread belief that young people would feel motivated to practice sports as an effect of the Olympic 'atmosphere'. Some researchers (Moore, 2013; Dolan et al., 2016) have shown that this magical atmosphere of optimism does exist, even if it is ephemeral. However, what the discourse around young people highlights is the expectation that people would act because they were moved by the 'inspiring' discourses of the IOC. This not only did not happen for London, but in fact never happens as demonstrated by research on sports inclusion (Hammond, 2016).

The interest in defending this agenda shows how important desire is for the legacy discourse. The way it is manifested in news articles is through the uses of words that, as I argued previously, contribute to a narrative that mimics the language use in utopian and dystopian fiction. Evidence of this was found by the literal reference to two dystopian films (The Omega Man and I am Legend) made by Carlos Brito in the article 'After four months, Olympic Park in Rio has signs of abandonment' (Brito, Globo, 2017). To describe the destruction of Rio's legacy, the author refers to it as a post-apocalyptic scene. The use of utopian and dystopian narratives, I argue, is part of an affective dispute.

6.6 An affective dispute

By looking at the distribution of sentiment labels assigned to news headlines (Figure 42), it becomes evident how neutral titles are scarce in the dataset in comparison to the positive and negative ones. These results show that news headlines reporting on legacy are very often composed of positive and negative words – mostly adjectives and adverbs. When applied to

news articles, sentiment analysis can do more than classify what is positive and what is negative. It helps in classifying 'how much' sentiment is present in the text. In other words, it measures the intensity of emotions in the articles, such as the use of 'huge' in 'Rio's Olympic legacy a "huge disappointment"' (Davies, BBC, 2017). And this alone provides further insight into the nature of the language used in the coverage of specific events. SA demonstrates that the coverage of legacy was far from neutral, sober or lacking in intensity. It was profoundly shaped by the use of strong qualifiers (e.g. abandoned, blessed, vandalised), emotion and affective disputes.

In Chapters 2.2 and 2.6 I touched on the temporality of legacy. The present chapter has shown how sentiment evolved over time, and how it is influenced by a sense of expectation before the event and surprise or frustration after the event. However, even when considerable transformations occur – as in the case of Rio when the positive expectation turned to very negative frustration – periods of positive media coverage can also be manipulated to support future discourses. This was the case of Boris Johnson's use of the 'successful' memory of the London Games to appease the fears of UK residents over Brexit. It was also the case in the Rio Mayor's campaigns to be elected Mayor of Rio de Janeiro again in 2020 and 2024. Despite the negative coverage of Rio's legacy, Paes accused others of interrupting the successful delivery of his plans and promised that, if reelected, he would fulfil everything that was promised before 2016. He was elected Mayor of Rio in 2020 and reelected in 2024. Eight years after the Rio Olympics, the event continued to impact the political discourse and the decisions made for, by and about the city. The level of positive sentiment that was detected in the news before the Rio Olympics served again to inspire and motivate people for the promise of changing the city for the better.

The way the sentiment in the news evolves overtime has political implications and serves the interests of multiple players involved in the Olympic legacy. In this sense, as well as serving politicians' careers, it also served, for example, the interests of London football clubs. The fear of white elephants endorsed by the Daily Mail and its many articles on how terrible an

athletics venue the Olympic Stadium would be, served the discourse that the best result for the Stadium would be for it to be handed over to a football club.

Positive and negative articles also serve the interests of the news outlets that publish them. In the case of Rio 2016, Globo presented the most positive coverage of the legacy of the Games, especially before and during the event (Figure 44). Globo was the official Brazilian broadcast of the Rio Games, and a positive expectation for the event would be beneficial to strengthen their audiences. In contrast, Folha de São Paulo, a news outlet based in São Paulo, was expected to receive few benefits from the Olympics, and presented a scarce and extremely negative coverage of the legacy.

In sum, in the same way positive coverage of the Games is important for the IOC and for the political actors involved in delivering the event, negative coverage can serve other interests. That is why I call it an 'affective dispute'. It is a discursive dispute to narrate the event and how it will be remembered. As this work resorted to Google to collect data, when googling 'London 2012 legacy' or 'Rio 2016 legacy', these are the stories that will tell the users of this widely used search engine what happened, whether it was successful or not, who was responsible for such success or failure, and what to expect from future mega events like the Olympiad.

However, it is important to highlight that the results and conclusions presented here are non-static. As discussed in section 3.6, this phenomenon must be read as a magmatic flow, in constant change. The memory of what London 2012 and Rio 2016 was is in constant dispute and will remain so. News stories continue to be published every year on these events and the narratives of legacy change. Therefore, this study is a snapshot of the current state-of-the-art, but the social phenomenon of Olympic legacy demands to be regularly revisited.

6.7 Conclusion

This chapter looked at sentiment outputs as a tool to analyse the media coverage of the Olympic legacies. According to the results, news headlines presented significantly more positive and negative than neutral labels, and tended to be more negative than their respective texts. When looking specifically at texts, neutrality increased significantly. These results are informative regarding the journalistic practices of using titles, especially online, to catch the attention of the audience.

By comparing the coverage of the two events, it is possible to conclude that London received more positive coverage while Rio coverage was more negative. While the Brazilian media has been less critical about London, the sentiments expressed by the British media about Rio were very negative. I have referred to this phenomenon as a utopian-dystopian dichotomy, where one event is represented as a 'huge disappointment' while the other one is a clear 'success', silencing or reducing the space in the media agenda for the nuances embedded in these complex media events.

By looking at the variation of sentiment over time, I identified that the British media was overall more sceptical about the legacy of the Games, while the Brazilian media was more optimistic. These results follow a similar pattern when compared with public opinion about the events published in surveys before and after each Olympics.

The intrinsic relationship between legacy and desire was explored, emphasising how it is manifested in the city as object of utopian desire and in the form of motivational and inspirational discourses. The use of SA was particularly important to unveil the numerical significance of subjective statements, the intensity of the language used, and to reveal what the positive and negative news headlines are, which were later inspected in more detail by close reading. To conclude, I defend the concept of legacy as the result of an affective dispute that serves the interest of multiple players involved in delivering the Games.

Chapter 7: Towards a conceptualisation of the narrated Olympic legacy

This chapter discusses the main findings of this thesis as well as its main contributions by returning to the concepts of Olympic legacy discussed in Chapter 2 in combination with the analysis developed in the subsequent chapters. As discussed in Chapter 2, there has been significant literature[84] produced in different fields of research such as sports studies, Olympic studies, urban studies, and the economic and social sciences that has attempted to conceptualise the Olympic legacy and its social, political and economic implications. What these studies do not cover in-depth, however, is the discursive dimension of legacy, constructed by and through the media coverage of Olympic events. This thesis, therefore, seeks to further understand how the concept of legacy is discursively manifested in news articles and what forms of narratives they constitute[85]. As pointed out in previous chapters, the reason why understanding this phenomenon becomes crucial is based on the understanding that news media collaborate in producing an understanding of the 'real world'. More specifically, the media documents and produces discourses that have the potential to become 'sites of reference' for the historical narratives that are formed.

7.1 A note on news articles as objects of study

There are some assumptions as to what news articles are, which need to be discussed here in order to clarify and conceptualise the object of this study. While presenting and discussing aspects of this thesis in conferences, I was frequently faced with questions or statements such as: (1) why are you studying sentiment in news articles? Aren't they supposed to be neutral?;

---

[84] Such as Cashman, 2005; Preuss, 2007; Gratton & Preuss, 2008; Poynter & MacRury, 2012; MacAloon, 2008; Girginov, 2018.

[85] Forms of narratives include positive, negative, utopian, dystopian, focused on tangible or intangible legacies, on national or international audiences, ones that privilege local communities or political and economic interests, among others.

or (2) news articles are easier to study using NLP because they use formal language in comparison to social media.

Although these questions did not come from experts in media studies, they represent more of a common sense approach arising from the expectations people have in relation to news articles and have served as a useful provocation for a reflection on the object of study. News articles constitute a much more diverse, noisy and complex genre than the formula assumed by readers as standard or the one found in newsroom manuals. Looking at the corpus of analysis, the previously discussed 'Tessa Jowell 'deserves an Olympic medal in utter madness' story (Samuel, Daily Mail, 2010) illustrates the use of figurative language (irony), for example. This is part of a phenomenon addressed by Lutpullaevna (2021), who identified the move of journalism away from the formal 'literary purity' towards a more emotional use of language, which is understood as more informed or influenced by societal changes. This is corroborated by Manjiri K. Carey, BBC News Labs editor, in an article published by the Reuters Institute for the Study of Journalism (Adami and Kahn, 2023). Carey argued that journalism has been moving away from traditional formats and structures such as the inverted pyramid[86]. For her, the audience is now demanding a more linear structure, which she describes as 'like a fairytale'.

Regarding people's perceptions concerning the lack of sentiment in news articles, the misunderstanding comes from the misconception of the term 'neutrality'. As mentioned in Chapter 6, neutrality in journalism refers to a desirable 'impartiality', which has a completely different meaning in sentiment analysis outputs. Moreover, the analysis of the news articles conducted in this study has shown that the boundaries between opinion and news can be hard to identify at first glance, or by using web scraping techniques. Many opinion articles have been published without a 'tag' identifying them as such. Figure 45 below shows how the information is presented.

---

[86] The inverted pyramid is a method to structure information on news articles, where the text should start by conveying the most important and relevant information in the first paragraphs and develop by focusing on specific and more detailed aspects.

Through the presence of Martin Samuel's name on the top of the page, accompanied by his picture, it becomes easier to identify that this is a column, where opinions are expected to have more space. However, by using the web scraping technique applied in this study, this information was missing. What we have instead is the news article's text, the URL, the name of the author and the headline. By looking at the URL 'https://www.dailymail.co.uk/sport/article-1252770/Martin-Samuel-Tessa-Jowell-deserves-Olympic-medal-utter-madness.html' it is not possible to identify whether it is an opinion article or a news piece.

Figure 45 – Screenshot of the news article 'Tessa Jowell deserves an Olympic medal in utter madness' published by the Daily Mail

The same happens in articles published by The Guardian (Figure 46). The presence of the author's picture helps to indicate the text may be an opinion piece, but this is part of the data that is not collected using the methodology applied in this study.

Figure 46 – Screenshot of the news article published by The Guardian, showing author's picture

In Folha, however, it is different. When publishing an opinion piece, it is found within an 'opinion' section on the website and, therefore, is recognisable by looking at the URL: 'https://www1.folha.uol.com.br/**opiniao**/2016/07/1793440-proteger-os-jogos.shtml'. It also includes the title *Opinião* on the top of the web page, besides the subtitle *Editorial* (Figure 47).

Figure 47 – Screenshot of an opinion article published by Folha de S. Paulo

Other important contextual features of specific articles have been difficult to identify because of the lack of signalling. This is the case of the previously mentioned article 'The Olympic legacy: creating a new community for London in Stratford' (Meredith, The Guardian, 2012), published in the news section. The article was written by the executive director of Triathlon Homes, the investor in properties in Stratford, and had no indication of essentially functioning as a piece of advertising.

As these limits become blurry and difficult to identify, it is important to consider that when talking about news articles as the objects of this study, they are not being considered as

objective, neutral and purely fact-based pieces. They are rather texts that sometimes include opinions, with some being signed by an author while others are not. The choice for not dividing articles into factual and opinion pieces, therefore, was not made solely based on the methodological challenge, but also because both genres work in combination to constitute media discourses. Having this 'ecosystem' in the background, this chapter will look at the different features of media coverage that constitute and collaborate in the conceptualisation of a 'narrated and mediated Olympic legacy'.

7.2 The media coverage of legacy is cyclical and historically informed

As discussed in section 2.2, the cyclical order in which events occur is part of the conceptualisation of media events, as described by Dayan and Katz (1992). When looking at the temporal aspects of the Olympic Games as a media event, it is possible to identify how previous experiences inform the media discourse of future Games. The modern Summer Olympic Games have been taking place every four years since 1896 (with few exceptions). Other major sporting events such as the Commonwealth Games, Winter Olympics and FIFA World Cup are also part of a cyclical agenda, in which journalists can look to the past to decide how the coverage of the next event will be framed.

As discussed in section 4.10, the news articles analysed tend to mention more than one previous or future host city, according to the network visualisations generated for the dataset composed of texts on the London Olympics. For Rio, however, the data does not present the same connections. Even though the networks for the Portuguese language did not produce as accurate results, a search for recent host cities such as London, Beijing and Tokyo[87] reveals that the number of mentions of those places in the dataset is relatively small.[88] According to the quantitative and qualitative analysis, it is possible to conclude that the Brazilian news

---

[87] In Portuguese: *Londres, Pequim* and *Tóquio*.

[88] Number of mentions in the dataset including Globo, Estadao and Folha de São Paulo (the three newspapers used for the qualitative analysis): London: 92; Beijing: 20; Tokyo: 48.

media coverage of Rio's legacy was focused primarily on local issues and the comparisons with other host cities were not as relevant as for the British media covering London.

This, however, does not mean that the Brazilian media discourse had no relation to previous coverage of Olympic legacies or even the legacy of other sporting events. *Copa Mundo* (World Cup) is the 37th most mentioned bigram in the corpus and was used as reference point of a recent Brazilian experience to inform the discourses on the expectations for the legacy of the Olympics.

The article 'Brazil agreed to pay the bill for the games, FGV executive criticises' (Amaral, Estadao, 2016), published one month before the opening ceremony of the Rio Olympics, presents an interview with Pedro Trengrouse, coordinator of the executive program for the improvement of Sports Management at Fundação Getúlio Vargas (FGV). Even though the interview was focused on the legacy of the Rio Olympics, Trengrouse was asked about the mistakes made by the Brazilian government with regards to the legacy of the 2014 FIFA World Cup. In his answer, he mentioned the lack of improvement in sports management in Brazil as well as issues with the management of stadiums after the event. He also drew a parallel with the World Cup to argue that the pessimism of the international media about the Olympics in Rio would not remain for long. For Trengrouse, as soon as the World Cup started in 2014, there was a shift in the international discourse, emphasising the 'magic' effects of the euphoria mentioned in previous chapters.

Although mentions of other Olympic host cities were less statistically significant in the Brazilian corpus, it is possible to see the reproduction of a logic, identified in the coverage of London, of expected overspending of public resources, the existence of white elephants (even though these words were not used as frequently in Portuguese as they were in English), and on the positive side, the expectations for the regeneration of the city.

Analysing the media coverage of different events across different national and regional contexts is, therefore, a useful mechanism for predicting the interest of the news media discourses and how the communication of public policies might be better planned to target

those specific and recurrent concerns. This could explain why Boris Johnson prioritised decisions about the future of the Olympic Stadium as a result of the past failings of Beijing in relation to their Olympic Stadium. Such a focus, however, may have been to the detriment of other legacies, which were less effectively communicated, and this is particularly true for intangible legacies.

7.3 A lack of space for the intangible

Girginov and Preuss (2022) provided a thorough literature review of the work to conceptualise the intangible legacy of the Olympics. The authors argued that, when thinking of the beneficiaries of this kind of legacy, it was necessary to move beyond an abstract idea of humanity to think of real individuals and how they would be impacted by such policies. Moreover, the authors highlighted the importance of producing qualitative measurement of intangible legacy, as it could not be quantified, and urged for it to be institutionalised to ensure its sustainability (Girginov and Preuss, 2022, p. 12).

As the study of the media coverage of London 2012 has shown, the main topics of interest in the media discourse were distant from official government communications published on the websites analysed. The interest in local issues, such as tangible school facilities and funding to support young people practising sports and the sustainable use of the Olympic Stadium, took precedence over the more intangible International Inspiration Programme. Besides the lack of a clear definition of the programme's main goals, the repeated slogan of 'inspiring' people to practise sports has been shown to have been a mistake from its conception. As discussed above (see Chapter 4.7), studies (Ken Green, 2012; Hammond, 2016) have demonstrated that the Olympic Games do not increase the number of residents in host cities taking part in sporting activities. Therefore, including this topic in the list of promises added an unrealistic objective to the legacy plan.

When looking at Rio 2016, however, it was found that discourses of intangible legacies were not even part of the - mostly informal - promises. From the beginning, the focus of Rio's

authorities as well as the local organising committee was on the tangible legacy of the Games. Part of this, as previously discussed, was grounded in the widespread idea that the Olympics were an opportunity for Rio's accelerated urban development and the expectation that cities in the Global South should prioritise specific local infrastructural issues and basic services rather than thinking and acting more globally, as London's international programme intended.

However a question this thesis can address is whether there is an effective way of communicating the intangible legacies. First, it is important to point out that intangible legacies were not a priority of the media coverage according to the data analysed for this study. For both the London and Rio Olympics, the main focus has been on the tangible legacies. This is to be expected based on how disruptive the construction work for the Games are and how they tend to become a visible monument of memory, reminding people of the event long after its conclusion..

As pointed out by Girginov and Preuss (2022), intangible legacies are key for the continuation of the Olympic Games. Popular foundational myths of the Olympic Games, such as the promotion – though sports – of a healthier, united (as suggested by the Olympic rings), peaceful world that embraces fair play, are part of what constitutes the magical atmosphere of this mega-event. However, projects such as the International Inspiration Programme also serve the interests of different groups who capitalise on the Olympic association and their shared values (Girginov, 2018, p. 83). This distinguishable status serves sponsors, the IOC, and governments particularly well. Therefore, the main strategy of the IOC and governments when communicating intangible legacies seems disassociated from the immediate socio-economic benefits expected from the Games – which are the prominent topics covered by the media – and should not be intertwined with other goals listed as promises.

This thesis has shown that promoting vague promises of intangible legacies was a mistake by London's organising committee. What is cited among the legacy promises needs to be less vague and measurable – quantitatively and qualitatively – to make sure that any contestation in the media can be responded to or defended by the IOC or local government. This way,

intangible legacies should be presented in more precise form and, perhaps, listed separately from the list of tangible promises, as they are essentially of a different nature. This makes particular sense if we consider how journalism tends to act as an institution responsible for monitoring the use of public money, and consequently prioritises tangible evidence of how this money is spent.

7.4 The media responds to (and constructs) the legacy

As pointed out by Felle, journalism plays an important role in democratic societies: that of watching governments and the public and private institutions that surround us. Because of its 'idealistic' independence from the executive, legislative and judiciary, news media is commonly referred to as the 'fourth estate' (Felle, 2015, p.2). This assumed role of acting as public watchdogs frames the media agenda and the ways in which news stories are reported. This characteristic of the work of journalists is particularly important for this study of the media coverage of the Olympic legacies because it helps to explain why facts related to the use of public money were often highlighted in the articles.

Given the release of a public document composed of promises in the case of the 2012 London Olympics, it was to be expected that journalists would look at those promises and respond to them when writing news stories. In the case of Rio, the work consisted more of gathering promises sporadically promoted by the government through official communication channels and in particular the Mayor's public speeches. In general, this thesis illustrates how the media coverage of the Olympic legacy is primarily a response of the media to the published or orally communicated legacy plans and how they are delivered.

Rio's authorities, as previously discussed, focused on the tangible legacy and avoided public documents containing promises. This could be viewed as a mechanism to avoid effective monitoring by the media and it tends to produce silences concerning other aspects of the legacy, such as the communities' claims for new jobs and new homes. Nevertheless, there are also ontological issues to be considered regarding how legacy is communicated in different

social, economic, cultural and linguistic contexts. Girginov alerts us to the fact that those differences regarding the diverse meanings of legacy (English) and *héritage* (French), as mentioned in Chapter 2, are not only linguistic but 'reflect different ontologies or worldviews about the nature of reality and legacy in particular' (Girginov, 2018, p. 198). On the one hand, it becomes difficult to look at both events and their media coverage through the same lenses, without considering their particularities; on the other hand, there is a global agenda that reverberates in both cases, such as the claims in relation to housing policies.

As providing new homes is typically not part of the promises for the Olympic legacy, this topic has become central for activists fighting against the use of public space for the construction of stadiums instead of benefiting people in need. Even though people being removed from areas for the construction of the Olympic Park has become cyclical in Olympic history, this does not receive as much space in the media coverage as the so-called Olympic legacy, as defined by the government and the Olympic authorities. The work of these activists, therefore, challenges the conceptualisation of Olympic legacy by widening its meanings and including their agendas within it.

Activists' discourses also challenge the hegemonic media discourse as they shed light on aspects of the Olympics that are not being covered, particularly those that show the potential of legacy to perpetuate inequality in society (Girginov, 2018, p. 206). However, besides the tension promoted by dissident voices, legacy as a topic to be covered by the media is still highly dependent on government investments and the reward citizens potentially receive as a material legacy.

By considering these aspects in combination, it is important to highlight that the media coverage of the Olympic legacy is informed by the legacy plans but it also constructs the legacy itself. It is based on this media coverage, analysed throughout this thesis, that aspects of these historical events are read and interpreted. The way the story is narrated matters to the IOC, to the governments, to the politicians, to the sponsors, to the communities and to all the groups involved in the making of the Olympics. This narration helps to define what is or is not a legacy, as well as how successful the delivery of each of the Games was. Moreover, it

shapes how future cities will plan their mega-events and which investments will be considered more cost-effective.[89] This study, therefore, contributes to an understanding of the past but also provides lenses to inform future mega-events.

7.5 Expectation, surprise and frustration set the tone of the legacy discourse

As discussed above, the understanding that news media not only narrates facts but works to 'organise' the world by constructing a fragment of reality is key for this study. Zizi Papacharissi (2015) argued that the media promotes a way of connecting people affectively to events that are physically distant from them. A fundamental point made by the author is that affect or emotions are not disconnected from rationality, as often depicted.

What this thesis has shown, based on the analysis of news articles published before, during and after each of the Olympic Games, is that there is an expectation before the Games about how the legacy will be delivered, which is followed by a sense of positive surprise or frustration regarding the materialised results. The sentiment analysis of news articles published before the Olympics has revealed that the British media coverage of the 2012 London Olympic Games was much more pessimistic about the event, while the Brazilian coverage of Rio was more optimistic. Afterwards, however, the British media had a positive reaction to how the Olympics and its legacies had been delivered smoothly as planned, while the Brazilian media demonstrated frustration.

The analysis of how the British media depicted the Rio Olympics and the Brazilian media depicted London, however, shows a different affective reaction. Before the event, the British media covered Rio in a more negative tone, expressing pessimism and doubt about Rio's capacity to deliver a mega-event. The Brazilian media, on the other hand, expressed an optimistic coverage of London. Due to the use of superlatives discussed in section 6.3 (e.g.

---

[89] By cost-effective, I mean in terms of communication. This study does not allow for an evaluation of the Olympic legacies, but rather provides an analysis of how legacy is communicated by the media.

abandoned, blessed, vandalised), I referred to the sentiment of the media discourses as often conforming to a utopian-dystopian dichotomy.

These highly affective responses to the legacy and how they change over time and space are, therefore, part of the conceptualisation of the Olympic legacy, based on how it is narrated by the media. In the legacy timeline, the concept of legacy is regularly reinterpreted, and new meanings as well as new sentiments are attributed to it many times over. Gratton and Preuss (2008, p.1924) have considered both 'positive and negative' structures as a result of the Games in their definition of Olympic legacy, despite the 'optimism bias' promoted by the IOC (Girginov, 2018, p.204). The conceptualisation of this thesis, looking at the media discourses, adds to this definition the unstable nature of the concept of legacy and how this has been shaped by expectations and been repeatedly reinterpreted later on. Beyond the change in the way legacy is framed over time, space is a key component of legacy discourse. Space shapes assumptions the media of a certain country has of another country's culture – such as the expectation for success or failure –, therefore impacting the journalistic narratives and promoting different affective responses in the media discourse.

7.6 The fear of failure shapes the legacy discourse

As discussed in the previous section, frustration can be the result of a negative legacy. But frustration can also be anticipated as fear of failure. In section 4.3, news articles covering Boris Johnson's concerns regarding the future of the 2012 London's Olympic stadium were analysed. By looking back at the Chinese Olympic Stadium for the 2008 Beijing Summer Olympics, the Bird's Nest, Johnson was reported by The Daily Mail to be fearing the existence of a white elephant as a result of the Games. In the London case, it would have

been particularly problematic as the venue was the most prominent building of the London Olympics[90].

The ghost of the white elephant surrounds the media coverage. Despite being a response to previous failures, on the one hand, it must also be read as an effective mechanism of political influence, on the other hand. The fear of building something that will be of no use supports arguments in favour of, for example, the privatisation of venues. News outlets, in this sense, must be seen as active actors and producers of political power, once they choose to publish stories that collaborate with this state of fear.[91]

When looking at Rio, it was found that there were recurrent mentions of authorities and the judiciary due to the impact of the Car Wash operation on the image of public institutions, which raised corruption schemes as part of this fear of failure. Furthermore, the Mayor of Rio, Eduardo Paes, emphasised in his interviews the amount of public money saved thanks to the investment of private companies in the Rio Olympics, as illustrated in the article 'Olympic legacy: Rio 2016 Tennis Centre goes to the private sector' (Globo, 2021b). In this article, Paes is quoted saying that 'it [Rio 2016 Olympic Games] was one of the cheapest Olympics in history, with the highest volume of private investment and there isn't a white elephant here'.

What this thesis has identified is that fear of failure is more than a result of those previously-mentioned cyclical and historically informed aspects of the media coverage of the Games. It is also a political argument, endorsed by news outlets, that supports policymakers' decisions, often prioritising the gains of private companies and politicians to the detriment of the public interest. It would be of particular importance, in this sense, for the audience to be better informed regarding the newspapers' corporate interests in order to be able to critically

---

[90] The Olympic Stadium in Stratford was the most prominent building especially because it was constructed for the Games. In other events, such as the Rio 2016 Olympics, the Olympic stadium did not have the same importance as it was a reused venue that was just adapted for the event.

[91] See section 4.4 where the article published by The Daily Mail is analysed.

analyse why these arguments are given space, and often with little attention to counter-arguments.

7.7 The legacy discourse is affectively mobilised by political figures

Sporting events 'constitute a key component of domestic and international development strategies' (Nauright, 2004, p.1325). As pointed out by Tomlinson and Young (2006), studying sporting spectacle as a media event cannot be dissociated from the study of power. As a form of cultural history, the legacy narrative is permeated by different 'conceptions, values and ideologies' (Tomlinson and Young, 2006, p.4). The Olympic Games is a prime example of sport as politics. Its global visibility and capacity for mobilising resources is not only a unique opportunity to accelerate urban development and to reimagine the host country's national and international identity; it also provides politicians with the opportunity to manage huge amounts of public money and be responsible for delivering an enduring legacy. The Olympic Games is therefore key for the study of public policies – planning, delivery and communication – and their relation to public opinion, and relatedly a countrys' national confidence.

Affect, in this context, is an essential fuel for political mobilisation. Papacharissi (2015, p.5) argues that 'without affect, feelings essentially do not "feel", for it is affect that provides the intensity with which we experience emotions like pain, joy, and love, and more important, the urgency to act upon those feelings'. Emotional and moralistic content is also known to spread further and faster on social media (Brady et al., 2017). An example of how affect is manifested in political narratives is the way in which Boris Johnson appropriated the legacy discourse in his favour, as discussed in section 4.11. Dolan et al. (2016) concluded that London citizens were significantly happier during the Games, but that this sentiment returned to normal levels in the year after.

Nostalgia was used by Boris Johnson to communicate a return to this happier state of feeling. In December 2019, as the UK Prime Minister, he tweeted his desire to host the football World

Cup in 2030. 'I want it to show our national confidence as we get Brexit done', wrote Johnson. This corroborates the argument that Boris Johnson used sporting events as discursive tools to communicate his political intentions and gain support for his political agenda. He was deliberately exploiting this sporting legacy in order to make an emotional connection with voters.

Like Boris Johnson, Eduardo Paes was also referred to as a 'legacy' of the Games. While the Games 'euphoria' is known to be ephemeral, benefiting from long-term positive responses to the legacy required communication strategies that disputed, after the Games, the concept of legacy.In February 2024, the Mayor of Rio, Eduardo Paes, and Brazilian president, Lula, participated in the opening ceremony of a public school in the Olympic Park in Rio. The school was built using parts of the venue used for martial arts competitions in 2016. Coincidently, both political figures were reappointed to the political roles they used to occupy in 2009, when Rio was elected the host of the 2016 Summer Games. In 2024, the Mayor of Rio and the Brazilian President relied on the Olympic legacy to dispute the memory of what the legacy of that event had been, as illustrated by Paes' speech: 'Never, in Olympic history, a Olympic venue became a school' (Boeckel, Globo, 2024). It is a dispute for who the authority responsible for delivering a successful legacy is – and also for what the legacy is – which contributes to their careers as being recognised as efficient administrators of the public goods.

7.8 Conclusion

This chapter discussed the most relevant findings of this research with regards to the conceptualisation of the Olympic legacy, based on the British and Brazilian media coverage of the 2012 London Olympic Games and the 2016 Rio Olympic Games.

The findings discussed allow us to develop a conceptualisation of the narrated Olympic legacy as: cyclical and historically informed; focused on tangible legacies; responsive to the 'official' legacy plans; affectively influenced by discourses that express expectation, surprise

and frustration; shaped by discourses of fear of failure, which can serve the interests of decision makers; and affectively mobilised by political figures.

Therefore, I further develop and add more complexity to Gratton and Preuss' (2008, p.1924) definition of Olympic legacy as 'planned and unplanned, positive and negative, intangible and tangible structures created through a sport event that remain after the event'. First, the unplanned legacy may be an object of dispute in the future, but it is the planned Olympic legacy that receives media attention due to the responsive characteristics of the media discourse.

Second, despite the IOC's 'positive bias' in their use of the concept of Olympic legacy, and the calls scholars have made to be aware of the existence of negative legacies, there is a 'negative bias' in the media discourse due to: (1) its 'watchdog mission': the expectation that the media should monitor the use of public money; (2) its political power in shaping narratives: negative discourses such as the fear of failure can serve as tools to defend private interests to the detriment of the public interest; and (3) its cyclical aspect: by looking at previous failures, there is an expectation of what might or might not work in the following events, which directs the media's attention.

Third, there is little space for intangible legacies in the media discourse. It is definitely present when political figures use affect to mobilise a positive discourse in favour of their own careers, especially when capitalising on the 'euphoria' of the Games. However, as something 'intangible' it is of a different nature to more tangible dimensions of Olympic legacy and does not receive attention from the media as part of the expectations for the legacy.

The fourth aspect relates to the temporality of legacy 'created through a sport event that remains after the event' (Gratton & Preuss, 2008, p.1924). This thesis has shown that a temporal definition of legacy is complex. It is difficult to define when it starts and when it ends, or if it even ends. In 2024, the year of the Paris Summer Games, politicians and the media in Rio were still discussing the legacy of the 2016 Olympics, for example, and

inaugurating schools as part of this legacy (Boeckel, Globo, 2024). It is also hard to define how long structures have to remain after the Games to be considered as legacy. As discussed, some 'successful' political careers were seen as a legacy of the Games by the media. Therefore, legacy narratives can be reimagined, resignified, destroyed and reconstructed and do not necessarily follow a strict timeline.

Rewriting Gratton and Preuss' (2008, p.1924) statement based on the analysis of the media narrative, it is possible to say that legacy is primarily composed of the officially planned tangible structures that are associated with the Olympic Games as a consequence of the event and is the product of a discursive dispute influenced by the media's negative bias.

Chapter 8: Conclusion

This thesis was driven by the following research question: 'what is the legacy of the London 2012 and Rio 2016 Olympic Games conceptualised through the British and Brazilian media coverage of these events?' As outlined in Chapter 1, its contribution is twofold: it contributes to the theoretical conceptualisation of Olympic legacy and it also offers a methodological pipeline for researchers interested in studying news media coverage of mega-events. Among the main objectives of this study were: (1) identifying in the news articles covering the legacy of the two events what aspects of Olympic legacy received attention; (2) identifying evidence of silences in the news media discourse by comparing them with activist and government discourses; (3) conceptualising Olympic legacy based on the aspects covered by the news media; and (4) providing researchers in the Digital Humanities with a methodological pipeline to collect and analyse multilingual news articles on a specific event by combining quantitative and qualitative approaches.

To respond to these questions and objectives, this study has made use of a combination of quantitative digital methods - with special focus on NLP techniques and network analysis - and qualitative methods - discourse analysis. The methodology was structured to allow for the identification of the lexicon of the media coverage of the legacy of the Games. It aimed at identifying the most relevant topics in the media discourse as well as those topics that, although important for the authorities or activist groups, did not receive significant attention from the hegemonic news media. For this purpose, a corpus containing news articles published by the most accessed British and Brazilian news websites was built using web scraping tools. Corpora of articles published by activist blogs and official government websites were also created for the purpose of comparison.

By analysing the collected data on the British media coverage of London 2012, it was found that special attention was given to the Olympic Stadium and discussions around its future as an Olympic legacy. Most articles debated whether the Olympic Stadium should have been

handed over to a football club or remained as an athletics venue. Boris Johnson's particular interest in the Olympic stadium shed further light on this topic, producing the most prominent controversy in the media narrative of London's legacy. This demonstrated the preference of the media for material legacy and how important material legacy is for bolstering politicians' reputations, mostly due to the fact that it represents an obvious and concrete outcome of the event.

The promise of engaging young people in sports activities was the most criticised legacy of London 2012 in the media discourse. Cuts to investment in sporting facilities at British schools were highlighted as one of the main causes for the decrease in sports participation. The drop in the participation of young people in sports activities shed light on the poor design of the legacy promises, where the London 2012 committee decided to include something that could not be guaranteed. In this case, the results are a warning for future committees to be mindful of the concrete possibilities of delivering what is promised.

Despite Boris Johnson having established himself as the leader of the planning for the Olympics, most of the negative aspects regarding the poor results of the plan for engaging young people were associated with Jeremy Hunt. In contrast, the legacy discourse strongly benefited Boris Johnson and his political career, which has been referred to as a legacy itself by the media. The political use of the Games – or of mega-events in general – is not new. Countries' main motivations behind hosting these events are often associated with the desire to showcase their success (or potential success). For politicians, it is always an opportunity to have their popularity boosted by the 'magical atmosphere of the Games'.

The use of NER and network graphs supported the conclusion that British news articles covering the London Olympics tended to cite previous Olympics to establish a comparison with London's performance for the 2012 Games. The references to other host cities shows how guaranteeing a successful delivery of the Olympic legacy is fundamental for the IOC's continued work. The fact that in recent years many cities have withdrawn their hosting bids

may be due to the fact that previous legacy-related experiences have been extremely negative. For the Games to continue, ensuring the delivery of the promised legacy is key.

The analysis of activists' blogs revealed that their concerns about the construction of affordable houses and the creation of new jobs received very little attention in the British media. The same applies to the struggle of those who had to move out of Stratford for the construction of the Olympic Park. On the one hand, these conclusions show that the struggle of local communities was not a relevant topic for the media, and that for neither politicians nor the IOC would it be an immediate issue if this remained silenced. On the other hand, this example shows how the mainstream media has been failing in its role of giving voice to the people and defending their interests. I would expect to see similar patterns in the future Olympics, and an analysis of how those voices will be represented would be beneficial for future studies.

Individual analysis of the coverage of each outlet was undertaken. The BBC presented the biggest diversity of topics in the coverage of London's legacy, covering the Olympic Park's construction, the access of young people to sports, physical education provision in schools, East London's redevelopment and political actions, with the latter mostly focused on Boris Johnson and Sebastian Coe. The Guardian focused on the political and economic effects of the legacy, looking primarily at the cost of the Games and the regeneration of East London. Public money, public spaces, public interest, public use and public spending are among the main topics discussed by this media outlet. The Daily Mail presented major coverage of the future of the Olympic Stadium. All other topics received much less attention. And The Telegraph was mainly focused on concerns about children and young people having access to sporting activities and the takeover of the Olympic Stadium by West Ham F.C.

The analysis of the individual news outlets strengthened the argument for the importance of a public media company and broadcast service like the BBC to ensure diversity in the media coverage. It also demonstrated how The Guardian tends to privilege debates that serve the public interest while The Daily Mail spent most of its time and space on mediating the fight

for the Olympic Stadium. The Telegraph was shown to be focused on local issues, which differs from the more global focus of the BBC and The Guardian.

By contrasting the media articles with government websites, it was found that the British government focused its communication strategies primarily on the International Inspiration programme, which was mostly composed of intangible promises. The government's official communications primarily focused on global issues, while activists focused on the needs of local communities. In contrast, the media was more concerned with the effects of legacy in London and the UK, rather than the impact it would have globally. It is comprehensible that London 2012 was aspiring to showcase the UK as a global leader. But what this analysis reveals is that this is not what the media believed was a priority for their audiences. In this sense, focusing on the International Inspiration programme was a mistake in the government's strategic communications.

Results of the sentiment analysis demonstrated the existence of a scepticism among the British media about the legacy of the London Olympics, which was evidenced by the negative sentiment of the news coverage before the event, followed by a positive (surprising effect) afterwards. The political implications of this sentiment transformation are important, having significantly contributed to Boris Johnson's discourse in which he presented himself as the 'father' of that success.

On the other hand, by analysing the Brazilian coverage of Rio 2016, it was found that there was an initial overexcitement being propagated in the media discourses, highly driven by the Mayor of Rio, about how the Olympics would promote radical and fast change in the city's infrastructure. The overexcitement was rapidly challenged by the amount of news articles shedding light on failures in delivering the legacy accordingly. Despite the potentially positive 'Green Games' slogan, the word 'green' was mostly associated in the news articles with the poor quality of the water in sports venues. Accidents with fires were also covered and contributed to the change in the optimistic coverage.

The media discourse about the Rio Olympics was found to have been significantly impacted by the political context that preceded the event. Police operations tracking corruption schemes in Brazil led to increased mentions in the media of authorities of the justice system and their attention to the decisions of policy makers. In this context, discourses in favour of privatising the legacy received special attention from the media. Most of the arguments worked to legitimise the investments in the Olympic Games based on private rather than public resources. This phenomenon raises questions about how much the discourses of corruption ended up serving to immobilise the State and favour the interests of the private sector. Once the State was seen as inefficient and corrupt by nature, handing over public infrastructure to be exploited by private initiatives became seen as a sensible option.

Eduardo Paes, the Mayor of Rio, centralised the legacy discourse in his hands (or, more appropriately, his mouth). Due to the lack of an official document stating the legacy plans, most of the promises were made by the Mayor in his public speeches and reverberated through the media. This concentrated power allowed the Mayor to adapt the legacy 'plans' according to his will throughout the event. Looking at government discourses in official websites has shown the disconnection between the Federal Brazilian Government and the local government of the city of Rio. The focus was given to programmes with the objective of supporting athletes financially. Centralising decisions is a commonplace if we look back at the analysis of London 2012 where Boris Johnson acted as the main figure behind the Olympic legacy. The Mayors of both cities found in the legacy narrative the opportunity to invest in their careers.

Data collected from activist blogs has shown how local issues in the city of Rio were connected with global issues and discussions about urban development. Blogs written in English raised concerns about the impact of the Olympics on residents of host cities, showing that what was happening in Rio was part of a global phenomenon and that the IOC should be involved in the solution. These global connections have already impacted future host cities, as the IOC demanded more planning and fewer removals of local residents for Tokyo 2020, for example.

The analysis of the discourses of each news outlet was also conducted for the Brazilian coverage of Rio 2016. Folha presented the most critical news coverage of Rio's Olympic legacy. The majority of the news articles collected for this research reported on negative aspects of the legacy. Among Folha's main topics were funding problems, the government's financial struggle and inefficiency in delivering the legacy. Globo presented the most diverse coverage, focusing not only on the problems but also echoing Eduardo Paes' expectations of an urban transformation as a result of the Games. Among Globo's most frequently occurring words was *atletas* (athletes). Athletes were mentioned by Globo mainly in the context of being a product of Rio's legacy, in terms of investment in their careers, or the emergence of new figures to be admired by the public. And Estadao's coverage presented a balance between the positive aspects of the games (sometimes based on promises before the games) and problems with concrete deliveries (negative aspects) after the Games.

The media landscape in Brazil is very different from the British one. This means that the lack of a strong public media outlet impacted access to a more diverse coverage. All the Brazilian media companies analysed in this thesis are dependent on sponsorship that often comes from banks and companies with an interest in the media agenda.

Returning to the analysis of the legacy discourse, I have argued that there was an aspect of *gambiarra* in the media narrative of Rio's legacy. Besides the fact that the event itself was disorganised in many ways, there was a lack of criticism of the media regarding this disorganisation. The Brazilian media covered Paes' public speeches as if they were sufficient evidence of a legacy promise to be accomplished. Temporal sentiment analysis combined with close reading revealed that the Brazilian media coverage of Rio was optimistic before the event and became more negative after. In this regard, Globo, in particular, promoted a very uncritical coverage of Rio's legacy prior to the event, which reveals an interest for building up positive expectations that would benefit the company, as a licensed host broadcast, with high audience rates.

When analysing the results of sentiment analysis, I conclude that legacy is intrinsically related to desire, which is manifested in the city and in the motivational and inspirational

discourses associated with the Games. This contributes to the understanding of legacy as an affective dispute, which serves the interest of several entities connected to the Games: such as the IOC, the politicians, different companies and sports organisations associated with the event, as well as the media itself.

These key findings listed above have contributed to the theoretical reflection on the conceptualisation of the narrated Olympic legacy based on the news media coverage. This way, I argue that the Olympic legacy is cyclical and historically informed, meaning that the media tends to look at previous events to address subsequent ones; is focused on tangible legacies, and therefore, intangible ones tend to be overlooked by the media; is structured as a response to the officially announced plans, whether formally (documented) or informally (political speeches) announced; is mobilised by affective discourses that express expectation, followed by responses of surprise and frustration; is shaped by discourses of fear of failure, which reverberate in the media and can serve the interests of decision makers; and is affectively mobilised by political figures, and in particular Mayors of host cities.

As a response to Gratton and Preuss' (2008, p.1924) definition of Olympic legacy as 'planned and unplanned, positive and negative, intangible and tangible structures created through a sport event that remain after the event', I offer my definition of Olympic legacy, as narrated by the media as being primarily composed of the officially planned tangible structures that are associated with the Olympic Games as a consequence of the event and is the product of a discursive dispute influenced by the media's negative bias.

These key findings contribute to the literature in the fields of sports, Olympic, urban and mega-events studies, by providing further insight into the definition of Olympic legacy, building particularly on the work of Gratton and Preuss (2008), Poynter and MacRury (2012), Cashman (2005), MacAloon (2008), and Girginov (2018). While definitions have been provided by the IOC, governments and scholars, this thesis offers a definition based on the media discourse, which constitutes an important part of how social phenomena are read and understood. Scholars in urban, sports and mega-events studies – such as such as Chappelet (2003); Hall (2006); Bairner & Molnar (2010); Dickson, Benson & Blackman (2011); and

Borges & Santos (2018) – have dedicated their research primarily to analysing historical documentation, the economics of the Olympics, the impact hosting such an event has in cities (public transportation, housing market, sports and leisure facilities), and the political implications of those decisions. This thesis fills the gap in scientific production on understandings of Olympic legacy as a concept constructed through the media discourse. Rather than focusing on how much money was spent (or invested), this thesis has looked at how the stories are told, that is how Olympic legacy is constituted through discourse.

The thesis also offers key contributions to the fields of Digital Humanities and media studies due to its methodological approach. It provides a framework to analyse news coverage of events by combining quantitative and qualitative methods, allowing researchers to handle massive amounts of data and retrieve significant insights from it. This work made use of several NLP techniques. Instead of word counts, the use of bigrams has been revealed to be more useful as it tends to provide greater context to the words. Obviously, some words do not need much contextualisation for an expert in the topic, such as 'ham' referring to 'West Ham'. However, other words, such as 'public' are better understood if analysed as a bigram containing, for example, 'public money', 'public interest', and so on.

The use of Named Entity Recognition (NER) has been shown to be useful for texts written in the English language. However, it presented significant challenges when applied to texts in Portuguese. Further work needs to be done in order to improve the capacities of these algorithms to recognise entities in languages other than English. I tested an alternative approach of combining Spacy with Wikifier to disambiguate entities and improve results, but it still failed for the Portuguese content (with some inconsistencies also in the English content).

Using network visualisations to illustrate entities from a narrative perspective has been shown to be promising. It helped to identify central actors in the media narrative in the English data. It also provided evidence of the cyclical aspect of the media coverage by showing how previous and future host cities were often mentioned together in the same article. It did not work well for Portuguese due to the previously mentioned issues with the NER. When these

methods failed, however, relying on tools such as Voyant tools or TIME, the latter developed for this project, has been useful.

As TIME has been mentioned, it is important to address the ephemerality of digital methods. TIME as a platform developed within the CLEOPATRA project suffers from maintenance issues, which have affected its usability for this research. Furthermore, the rapid release of new technologies makes some of the methodological approaches of this thesis unreproducible in the exact way they were designed. Over the last five years of development of this study, tools such as Boilerpipe were modified in such a way that they do not need JAVA intersection anymore and work fully in Python. New web scrapers were released, such as Trafilatura, which have shown the potential to extract textual information from websites with more precision − without adding too much irrelevant content −, and dealing better with diverse webpages' infrastructure than the tools used in this thesis. In the meantime, a revolutionary tool (Chat GPT) has been made publicly available, allowing not only for more efficient results regarding the identification of sentiment in texts, but also popularising the potential use of digital methods and AI in general. Results derived from the use of generative AI, however, are typically not transparent and reproducible in the same way as much of the research presented here.

Despite these changes, the main methodological contributions of this thesis lie in the workflow and how quantitative and qualitative methods were combined. The first works as a means to explore the content at scale and generate insights from the data, and the second operates as a method to contextualise these insights to produce knowledge. Moreover, this thesis has dedicated significant space to reflect on the use of these methods and their potentials and limitations. The use of explainable AI tools, for example, show a potentially useful way of interrogating algorithms to improve researchers' development and use of analytical tools. Faced with the emergence and popularisation of generative AI, understanding the results regenerated by these tools tends to become harder and harder. This is a particular challenge for research in the humanities and social sciences, as answering questions such as 'on which data has this module been trained?' or 'how did this module

produce these outputs based on my query?' will become more and more complex. As a guard against the ephemerality of the digital, and positioning this thesis as a legacy in itself, this work has also archived all the websites used as primary sources in the Internet Archive.

Regarding Sentiment Analysis (SA), and the profound reflection on its use for this study, it was found that SA of news headlines is more useful than sentiment analysis of the articles' overall body text. In order to make the analysis of the body text more useful, it would have been necessary to test, for example, the use of targeted sentiment analysis (entity level) or a similar approach. Sentiment classification of the Portuguese language also performed poorly. Translation, however, has been shown to be an efficient way of handling the lack of more suitable algorithms.

Sentiment classification still presents significant inconsistencies, even in English. The limitations of the method, such as the struggle with sarcasm and domain-specificity, were discussed in this thesis and should be accounted for when using its results to form general conclusions. Ultimately, interrogating sentiment algorithms using explainable AI tools such as SHAP has been shown to be useful in identifying some of their limitations.

Overall, analysing sentiment in news articles has been shown to be an effective way of studying the coverage of events. It provided a general understanding of how positively or negatively the event was narrated by the media, and insights about how this changed were relevant for understanding the specific dynamics of expectation followed by surprise or frustration in the Olympic legacy context. Future work could consider combining SA with emotion analysis in order to deepen the results and further contextualise different characteristics of the discourse.

Besides its theoretical and methodological contributions, the findings of this thesis have potential practical implications for communication strategies adopted by governments and event organisers, whether the IOC or organisers of any other mega-event. The analysis of the media coverage of London 2012 and Rio 2016 has shown how other mega-events are impacted by the legacy of the Olympics and by the way the Olympic Games are narrated,

including but not limited to the FIFA World Cup. This thesis, therefore, provides insights into aspects of the media coverage that have been shown to be recurrent, allowing for more effective planning in relation to how promises are to be communicated and how the organisers' agendas could be better operationalised. The thesis also serves to aid wider news audiences in understanding how legacy is historically narrated, providing tools for interpreting news articles and information on future events covered by the media.

This study has nevertheless faced many limitations. Besides the previously mentioned issues with algorithms for analysing data in Portuguese language, the lack of a Brazilian web archive for retrieving data and the concentration of media ownership in Brazil have limited the corpus construction and the possibility of comparing and contrasting results obtained from articles covering the London 2012 Games. Those limitations, however, have become an important contribution of this thesis with regards to multilinguality. Working with data in English and Portuguese has not only challenged the author in technical terms, but also in relation to the interpretative work. Despite working with machine translation, the inconsistencies in the number and quality of sources required the search of extra-corpus elements that could guide the interpretative work. While the scientific literature looking at London 2012 is vast, the lack of published works looking at Rio 2016 made it necessary to look for works on other events occurring in Brazil, such as the FIFA World Cup 2014, as well as events occurring abroad. It is also important to point out how specific media landscapes affect archival records. The concentration of media ownership in Brazil means less diversity and fewer resources to interpret events in the country.

This study has required significant time to be dedicated to learning digital methods. This process of learning by doing often results in choices that might not have been optimal. An example of this is the way the data was stored as .txt files. Were this research to be undertaken again, I would rather have stored the data as .json, including more metadata that would better facilitate the analysis.

This thesis did not include analysis of how the media narratives spread on social media and how they were resignified when entering those spaces. This limitation should be addressed in

future work, considering that social media is one of the most important mediums for the circulation of news articles, and that its affordances promote ways of interacting with the content that can also shape how it will be interpreted by the audience.

More research on the media coverage of the Olympic legacies is also necessary. Future work could potentially expand the corpus by including other Olympic Games, such as Tokyo 2020 and Paris 2024. Tokyo would be particularly interesting, as it was delivered after a delay in 2021 and under unprecedented circumstances due to the Covid-19 pandemic. It would also be beneficial to look more closely at the discourses of political figures and how they are affectively mobilised in different contexts. Identifying these strategies more comprehensively would make a valuable contribution to understanding how these mega-events serve political careers. Research dedicated to the media coverage of the legacy of the Paralympic Games would also be important, especially to identify how or if it differs precisely from the general Olympics legacy discourse.

The pipeline applied here could, therefore, be replicated to study the coverage of future Olympic Games, other sporting or mega-events, or even events of a different nature. The specific characteristics of these events, however, must be taken into consideration. The legacy of the Olympics, for example, is a controversial topic that produces several counter-discourses, including positive and negative articles, which ensured that SA was a particularly insightful methodological approach.

As a thesis in Digital Humanities, constructing this work has been challenging, as it draws on and seeks to offer contributions to several disciplines. The specific value of the Digital Humanities was, however, in the choice of lenses I would use to interrogate my object of study. The decision to use digital methods to explore the data has shaped where and how I would look and which aspects would be deemed relevant to be studied. These choices overlook some specificities – such as topics that did not receive statistically significant media coverage and, therefore, could not be listed among the most occurring n-grams –, which I have tried to take account of by using contrasting sources and combining these methods with close reading. Yet, overall, this methodological approach presented an innovative approach to

the study of media events, and one that allows for consideration of a wider number of sources and a more comprehensive perspective on the most relevant topics covered by the media across two different national contexts.

Therefore, this thesis has not only contributed to the knowledge of how Olympic legacies come about but has also reinforced the potential of digital methods for answering questions that are important to the humanities and social sciences more broadly, using them not just as a tool, but as an alternative mechanism to reflect on social phenomena.

References

Primary sources

AGLO (2017a). Legay Plan Rio 2016. Retrieved August 7, 2019, from
http://www.rededoesporte.gov.br//pt-br/noticias/aglo-apresenta-plano-de-legado-das-instalaco
es-olimpicas-e-matriz-de-responsabilidades-dos-jogos-rio-2016/plano-de-legado-das-instalac
oes-olimpicas. Archived on:
<https://web.archive.org/web/20220726080450/http://www.rededoesporte.gov.br//pt-br/notici
as/aglo-apresenta-plano-de-legado-das-instalacoes-olimpicas-e-matriz-de-responsabilidades-d
os-jogos-rio-2016/plano-de-legado-das-instalacoes-olimpicas>. Archive date: 26th July 2022.

AGLO (2017b). Autoridade de Governança do Legado Olímpico Cartilha. Ministério do
Esporte. Retrieved April 28, 2021, from
http://www.rededoesporte.gov.br//pt-br/legado/cartilha_AGLO_web.pdf.

Adami, M. and Kahn, G. (2023). What we learnt at our 40-year Fellowship reunion about the
future of news. Available on:
<https://reutersinstitute.politics.ox.ac.uk/news/what-we-learnt-our-40-year-fellowship-reunio
n-about-future-news>. Archived on:
<https://web.archive.org/web/20230911161427/https://reutersinstitute.politics.ox.ac.uk/news/
what-we-learnt-our-40-year-fellowship-reunion-about-future-news>. Archive date: 11th
September 2023.

Amaral, L. (2016). Brazil agreed to pay the bill for the games, FGV executive criticises.
Estadao. Available on:
<https://www.estadao.com.br/esportes/olimpiadas/brasil-aceitou-pagar-a-conta-dos-jogos-crit
ica-executivo-da-fgv/>. Archived on:
<https://web.archive.org/web/20231004162538/https://www.estadao.com.br/esportes/olimpia
das/brasil-aceitou-pagar-a-conta-dos-jogos-critica-executivo-da-fgv/>. Archive date: 4th
October 2023.

Anderson, R. (2012). Olympic legacy: White elephant or economic viagra?. BBC. Available
on: <https://www.bbc.com/news/business-19191758>. Archived on:
<https://web.archive.org/web/20230918190006/https://www.bbc.com/news/business-191917
58>. Archive date: 18th September 2023.

Arbex, T. (2016). Olympics raises Paes as candidate for governor of Rio in 2018. Folha.
Available on:
<http://www1.folha.uol.com.br/esporte/olimpiada-no-rio/2016/08/1805933-olimpiada-alca-p
aes-a-candidato-ao-governo-do-rio-em-2018.shtml>. Archived on:
<https://web.archive.org/web/20230918182437/https://www1.folha.uol.com.br/esporte/olimp

iada-no-rio/2016/08/1805933-olimpiada-alca-paes-a-candidato-ao-governo-do-rio-em-2018.s
html>. Archive date: 18th September 2023.

Azevedo, E. (2019). Seeds planted by athletes at the opening of Rio 2016 became a forest.
Globo. Available on:
<https://oglobo.globo.com/esportes/sementes-plantadas-por-atletas-na-abertura-da-rio-2016-v
iraram-floresta-veja-fotos-25123525>. Archived on:
<https://web.archive.org/web/20230918181021/https://oglobo.globo.com/esportes/sementes-
plantadas-por-atletas-na-abertura-da-rio-2016-viraram-floresta-veja-fotos-25123525>.
Archive date: 18th September 2023.

Barros, T. (2016). Olympic legacy: emergence of Idols is an incentive to practice sports.
Globo. Available on:
<http://globoesporte.globo.com/eu-atleta/saude/noticia/2016/08/legado-olimpico-surgimento-
de-idolos-e-estimulo-pratica-esportiva.html>. Archived on:
<https://web.archive.org/web/20230918185616/https://ge.globo.com/eu-atleta/saude/noticia/2
016/08/legado-olimpico-surgimento-de-idolos-e-estimulo-pratica-esportiva.html>. Archive
date: 18th September 2023.

BBC (2004). UK urged to prepare for Olympics. Available on:
<http://news.bbc.co.uk/2/hi/uk_news/3949875.stm>. Archived on:
<https://web.archive.org/web/20230918174613/http://news.bbc.co.uk/2/hi/uk_news/3949875.
stm,>. Archive date: 18th September 2023.

BBC (2010). West Ham and Newham Council eye bid for Olympic Stadium. Available on:
<http://news.bbc.co.uk/sport2/hi/football/teams/w/west_ham_utd/8583608.stm>. Archived
on:
<https://web.archive.org/web/20230918171343/http://news.bbc.co.uk/sport2/hi/football/team
s/w/west_ham_utd/8583608.stm>. Archive date: 18th September 2023.

BBC (2010b). West Ham's Olympic stadium plan angers Kelly Sotherton\n
. Available on:
<http://news.bbc.co.uk/sport1/hi/football/teams/w/west_ham_utd/8533774.stm>. Archived
on:
<https://web.archive.org/web/20230918171320/http://news.bbc.co.uk/sport1/hi/football/team
s/w/west_ham_utd/8533774.stm>. Archive date: 18th September 2023.

BBC (2011). Tottenham would demolish Olympic Stadium after Games. Available on:
<https://www.bbc.com/news/uk-england-london-12173663>. Archived on:
<https://web.archive.org/web/20230918173734/https://www.bbc.com/news/uk-england-londo
n-12173663>. Archive date: 18th September 2023.

BBC (2012). Boris Johnson urges two hours of PE a day. Available on:
<https://www.bbc.com/news/uk-politics-19192153>. Archived on:
<https://web.archive.org/web/20230918174919/https://www.bbc.com/news/uk-politics-19192
153>. Archive date: 18th September 2023.

BBC (2013). Jerome Valcke: Fifa chief says too much democracy can be hindrance.
Available on: <https://www.bbc.com/sport/football/22288688>. Archived on:
<https://web.archive.org/web/20230918180231/https://www.bbc.com/sport/football/2228868
8>. Archive date: 18th September 2023.

BBC (2015). Key London Olympics legacy 'a failure', says Tessa Jowell. Available on:
<https://www.bbc.com/news/uk-33404776>. Archived on:
<https://web.archive.org/web/20230918174255/https://www.bbc.com/news/uk-33404776>.
Archive date: 18th September 2023.

BBC (2017). Olympic legacy: did £1bn after 2012 get any more people doing sport?.
Available on: <https://www.bbc.com/news/uk-england-40817063>. Archived on:
<https://web.archive.org/web/20230918174323/https://www.bbc.com/news/uk-england-4081
7063>. Archive date: 18th September 2023.

BBC (2022). Licence fee and funding. Available on:
<https://www.bbc.com/aboutthebbc/governance/licencefee>. Archived on:
<https://web.archive.org/web/20230907185055/https://www.bbc.com/aboutthebbc/governanc
e/licencefee>. Archive date: 7th September 2023.

BBC (2023). Brazil media guide. Available on:
<https://www.bbc.co.uk/news/world-latin-america-18912764>. Archived on:
<https://web.archive.org/web/20230918185347/https://www.bbc.com/news/world-latin-ameri
ca-18912764>. Archive date: 18th September 2023.

Boeckel, C. (2024). Lula and Paes inaugurate a school named after Isabel Salgado in the
Olympic Park. Globo. Available on:
<https://g1.globo.com/rj/rio-de-janeiro/noticia/2024/02/07/lula-e-paes-inauguram-no-parque-
olimpico-escola-com-o-nome-de-isabel-salgado.ghtml>. Archived on:
<https://web.archive.org/web/20240215034849/https://g1.globo.com/rj/rio-de-janeiro/noticia/
2024/02/07/lula-e-paes-inauguram-no-parque-olimpico-escola-com-o-nome-de-isabel-salgad
o.ghtml>. Archive date: 15th February 2024.

Bond, D. (2013). London 2012: Has Britain made the most of its Olympic legacy?. BBC.
Available on: <https://www.bbc.co.uk/sport/olympics/23467773>. Archived on:
<https://web.archive.org/web/20230918174714/https://www.bbc.com/sport/olympics/234677
73>. Archive date: 18th September 2023.

Bortoloti, M. (2016). Why is Guanabara Bay still polluted during the Olympic Games?.
Globo. Available on:
<https://epoca.oglobo.globo.com/esporte/olimpiadas/noticia/2016/08/por-que-baia-de-guanab
ara-continua-poluida-nos-jogos-olimpicos.html>. Archived on:
<https://web.archive.org/web/20230918180855/https://epoca.oglobo.globo.com/esporte/olim
piadas/noticia/2016/08/por-que-baia-de-guanabara-continua-poluida-nos-jogos-olimpicos.ht
ml>. Archive date: 18th September 2023.

Boulos, G. (2016). Olympic village and the Vila Autodromo. Folha. Available on: <http://www1.folha.uol.com.br/colunas/guilhermeboulos/2016/07/1796295-a-vila-olimpica-e-a-vila-autodromo.shtml>. Archived on: <https://web.archive.org/web/20230918185149/https://www1.folha.uol.com.br/paywall/login.shtml?https://www1.folha.uol.com.br/colunas/guilhermeboulos/2016/07/1796295-a-vila-olimpica-e-a-vila-autodromo.shtml>. Archive date: 18th September 2023.

Brain, J. (2012). Boris Johnson announces cycling legacy plans. BBC. Available on: <https://www.test.bbc.com/news/uk-19206162>. Archived on: <https://web.archive.org/web/20230918175037/https://www.test.bbc.com/news/av/uk-19206162>. Archive date: 18th September 2023.

Brito, C. (2017). After four months, Olympic Park in Rio has signs of abandonment. Globo. Available on: <https://g1.globo.com/rio-de-janeiro/noticia/apos-quatro-meses-parque-olimpico-no-rio-tem-sinais-de-abandono.ghtml>. Archived on: <https://web.archive.org/web/20230918181828/https://g1.globo.com/rio-de-janeiro/noticia/apos-quatro-meses-parque-olimpico-no-rio-tem-sinais-de-abandono.ghtml>. Archive date: 18th September 2023.

Brown, O. (2017). Rio's Olympic legacy lies in ruins - even if the IOC continues to insist otherwise. Telegraph. Available on: <https://www.telegraph.co.uk/olympics/2017/03/16/rios-olympic-legacy-lies-ruins-even-ioc-continues-insist-otherwise/>. Archived on: <https://web.archive.org/web/20230918190801/https://www.telegraph.co.uk/olympics/2017/03/16/rios-olympic-legacy-lies-ruins-even-ioc-continues-insist-otherwise/>. Archive date: 18th September 2023.

Buzzani, A. (2022). Olympic Legacy: Rio begins work to transform Arena do Futuro into schools. Band. Available on: <https://www.band.uol.com.br/bandnews-fm/rio-de-janeiro/noticias/legado-olimpico-rio-inicia-obras-para-transformar-arena-do-futuro-em-escolas-16501408>. Archived on: <https://web.archive.org/web/20230918182622/https://www.band.uol.com.br/bandnews-fm/rio-de-janeiro/noticias/legado-olimpico-rio-inicia-obras-para-transformar-arena-do-futuro-em-escolas-16501408>. Archive date: 18th September 2023.

Carvalho, L. (2016). Eduardo 'Pereira' Paes, the Brazilian Olympics' Haussmann. Estadao. Available on: <https://brasil.estadao.com.br/blogs/tudo-em-debate/eduardo-pereira-paes-o-haussman-da-olimpiada-no-brasil/>. Archived on: <https://web.archive.org/web/20211020021730/https://brasil.estadao.com.br/blogs/tudo-em-debate/eduardo-pereira-paes-o-haussman-da-olimpiada-no-brasil/>. Archive date: 20th October 2021.

Coelho, H. (2013). Eduardo Paes does not reveal Olympic budget and shoots against FIFA. Globo. Available on:

<http://g1.globo.com/rio-de-janeiro/noticia/2013/08/eduardo-paes-nao-revela-orcamento-oli mpico-e-dispara-contra-fifa.html>. Archived on:
<https://web.archive.org/web/20230918182128/https://g1.globo.com/rio-de-janeiro/noticia/20 13/08/eduardo-paes-nao-revela-orcamento-olimpico-e-dispara-contra-fifa.html>. Archive date: 18th September 2023.

Conn, D. (2006). No cut-price Olympic legacy for football's fat cats. The Guardian. Available on: <https://www.theguardian.com/sport/2006/oct/11/Olympics2012.politics>. Archived on: <https://web.archive.org/web/20230918170210/https://www.theguardian.com/sport/2006/oct/ 11/Olympics2012.politics>. Archive date: 18th September 2023.

Crerar, P. (2008). Boris Johnson promises 'tens of millions' to realise 2012 vision of a British Olympics legacy in 2012. Daily Mail. Available on:
<https://www.dailymail.co.uk/news/article-1049276/Boris-Johnson-promises-tens-millions-re alise-2012-vision-British-Olympic-legacy-2012.html>. Archived on:
<https://web.archive.org/web/20230918174110/https://www.dailymail.co.uk/news/article-104 9276/Boris-Johnson-promises-tens-millions-realise-2012-vision-British-Olympic-legacy-201 2.html>. Archive date: 18th September 2023.

Daily Mail (2008). Spurs and West Ham bid for new homes as 2012 venues up for grabs. Available on:
<https://www.dailymail.co.uk/sport/football/article-1052688/Spurs-West-Ham-bid-new-home s-2012-venues-grabs.html>. Archived on:
<https://web.archive.org/web/20230918170317/https://www.dailymail.co.uk/sport/football/ar ticle-1052688/Spurs-West-Ham-bid-new-homes-2012-venues-grabs.html>. Archive date: 18th September 2023.

Department for culture, media and sport (2008). *Before, during and after: Making the most of the London 2012 Games*. 84.

Diario do Rio (2022). Rio City Hall begins dismantling the Arena of the Future, in the Olympic Park. Available on:
<https://diariodorio.com/prefeitura-do-rio-inicia-a-desmontagem-da-arena-do-futuro-no-parq ue-olimpico/>. Archived on:
<https://web.archive.org/web/20230918185840/https://diariodorio.com/prefeitura-do-rio-inici a-a-desmontagem-da-arena-do-futuro-no-parque-olimpico/>. Archive date: 18th September 2023.

Donahue, B. (2020). The price of Gold: one village's resistance sparked a global question: should the Games be abolished?. Rio on Watch. Available on:
<https://rioonwatch.org.br/?p=48779>. Archived on:
<https://web.archive.org/web/20230918185002/https://rioonwatch.org.br/?p=48779>. Archive date: 18th September 2023.

Ecorys (2014). Final Evaluation of the International Inspiration Programme. Available on: <https://www.britishcouncil.org/sites/default/files/ecorys_international_inspiration_final_revi ew_2014_1.pdf>. Archived on:

<https://web.archive.org/web/20231014194208/https://www.britishcouncil.org/sites/default/files/ecorys_international_inspiration_final_review_2014_1.pdf>. Archive date: 14th October 2023.

Engel, M. (2016). The trouble with the Olympics Games. The Guardian. Available on: <https://www.theguardian.com/sport/2016/jul/29/the-trouble-with-the-olympic-games>. Archived on: <https://web.archive.org/web/20230918174757/https://www.theguardian.com/sport/2016/jul/29/the-trouble-with-the-olympic-games>. Archive date: 18th September 2023.

Estadao (2011). London admits struggle to get Olympic legacy. Available on: <https://esportes.estadao.com.br/noticias/geral,londres-admite-dificuldade-para-obter-legado-olimpico,775569>. Archived on: <https://web.archive.org/web/20230918191119/https://www.estadao.com.br/esportes/londres-admite-dificuldade-para-obter-legado-olimpico/>. Archive date: 18th September 2023.

Estadao (2012). Cameron Prescribes Hard Work for Olympic Success. Available on: <https://www.estadao.com.br/esportes/cameron-receita-trabalho-duro-para-ter-sucesso-olimpico/>. Archived on: <https://web.archive.org/web/20230918191037/https://www.estadao.com.br/esportes/cameron-receita-trabalho-duro-para-ter-sucesso-olimpico/>. Archive date: 18th September 2023.

Estadao (2012b). Rio suffers from 'violence, bribery, favelas and decrepit Maracanã', says British newspaper. Available on: <https://politica.estadao.com.br/noticias/geral,rio-sofre-com-violencia-suborno-favelas-e-maracana-decrepito-diz-jornal-britanico,939529>. Archived on: <https://web.archive.org/web/20230918185651/https://www.estadao.com.br/politica/rio-sofre-com-violencia-suborno-favelas-e-maracana-decrepito-diz-jornal-britanico/>. Archive date: 18th September 2023.

Estadao (2013). After Games, Handball arena will be transformed into four schools. Available on: <https://esportes.estadao.com.br/noticias/geral,apos-jogos-do-rio-de-2016-arena-de-handebol-vai-virar-quatro-escolas,1096742>. Archived on: <https://web.archive.org/web/20230918180457/https://www.estadao.com.br/esportes/apos-jogos-do-rio-de-2016-arena-de-handebol-vai-virar-quatro-escolas/>. Archive date: 18th September 2023.

Estadao (2015). Rio focuses on efficiency for the Olympic Games. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,rio-aposta-na-eficiencia-dos-projetos-para-os-jogos-olimpicos,1787122>. Archived on: <https://web.archive.org/web/20230918180547/https://www.estadao.com.br/esportes/olimpiadas/rio-aposta-na-eficiencia-dos-projetos-para-os-jogos-olimpicos/>. Archive date: 18th September 2023.

Estadao (2016). Problems in construction threatens city's Olympic legacy. Available on: <https://brasil.estadao.com.br/noticias/rio-de-janeiro,problemas-em-obras-ja-ameacam-o-lega

do-olimpico-da-cidade,10000027731>. Archived on: <https://web.archive.org/web/20230918180627/https://www.estadao.com.br/brasil/rio-de-jan eiro/problemas-em-obras-ja-ameacam-o-legado-olimpico-da-cidade/>. Archive date: 18th September 2023.

Estadao (2017). Former Sports Minister was fined for lack of legacy after the Rio 2016 Games. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,ex-ministro-do-esporte-e-multado- por-falta-de-legado-apos-jogos-do-rio-2016,70001728192>. Archived on: <https://web.archive.org/web/20230918184717/https://www.estadao.com.br/esportes/jogos-ol impicos/ex-ministro-do-esporte-e-multado-por-falta-de-legado-apos-jogos-do-rio-2016/>. Archive date: 18th September 2023.

Estadao (2017). After five months, Olympic legacy presents more negative points than positive. *Estadao*. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,apos-cinco-meses-legado-olimpico -apresenta-mais-pontos-negativos-do-que-positivos,70001656292>. Archived on: <https://web.archive.org/web/20220228152300/https://esportes.estadao.com.br/noticias/jogos -olimpicos,apos-cinco-meses-legado-olimpico-apresenta-mais-pontos-negativos-do-que-posit ivos,70001656292>. Accessed on 15th January, 2022.

Estadao (2017b). Rio suffers from expensive and empty arenas after the Olympics. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,rio-sofre-com-arenas-caras-e-vazia s-um-ano-apos-os-jogos-olimpicos,70001923539>. Archived on: <https://web.archive.org/web/20230918181730/https://www.estadao.com.br/esportes/olimpia das/rio-sofre-com-arenas-caras-e-vazias-um-ano-apos-os-jogos-olimpicos/>. Archive date: 18th September 2023.

Estadao (2018). Barra Olympic Park will have a national sports museum. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,parque-olimpico-da-barra-tera-mus eu-do-esporte-nacional,70002168369>. Archived on: <https://web.archive.org/web/20230918185720/https://www.estadao.com.br/esportes/olimpia das/parque-olimpico-da-barra-tera-museu-do-esporte-nacional/>. Archive date: 18th September 2023.

Estadao (2018b). Finally, Olympic legacy for Brazilian sport arrives in Deodoro. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,enfim-legado-olimpico-para-o-espo rte-brasileiro-chega-em-deodoro,70002168365>. Archived on: <https://web.archive.org/web/20230918185745/https://www.estadao.com.br/esportes/olimpia das/enfim-legado-olimpico-para-o-esporte-brasileiro-chega-em-deodoro/>. Archive date: 18th September 2023.

Estadao (2018c). Rio's victims of violence try to reinvent themselves in canoeing. Available on:

<https://esportes.estadao.com.br/noticias/geral,vitimas-da-violencia-no-rio-tentam-se-reinven
tar-na-canoagem,70002230765>. Archived on:
<https://web.archive.org/web/20230918184232/https://www.estadao.com.br/esportes/vitimas-
da-violencia-no-rio-tentam-se-reinventar-na-canoagem/>. Archive date: 18th September
2023.

Estadao (2018d). Union will spend R$ 35 million to keep the Olympic Park functioning.
Available on:
<https://esportes.estadao.com.br/noticias/jogos-olimpicos,uniao-vai-gastar-r-35-milhoes-para
-manter-funcionamento-do-parque-olimpico,70002149738>. Archived on:
<https://web.archive.org/web/20230918181754/https://www.estadao.com.br/esportes/olimpia
das/uniao-vai-gastar-r-35-milhoes-para-manter-funcionamento-do-parque-olimpico/>.
Archive date: 18th September 2023.

Filipo, L. (2015). Mayor office launches legacy plan and exposes battle with Ministry of
Sports. Globo. Available on:
<https://globoesporte.globo.com/olimpiadas/noticia/2015/07/plano-de-legado-olimpico-expo
e-rusga-entre-prefeitura-e-ministerio-do-esporte.html>. Archived on:
<https://web.archive.org/web/20230918180310/https://ge.globo.com/olimpiadas/noticia/2015
/07/plano-de-legado-olimpico-expoe-rusga-entre-prefeitura-e-ministerio-do-esporte.html>.
Archive date: 18th September 2023.

Folha (2012). Final cost of the Games must burst. Available on:
<https://www1.folha.uol.com.br/fsp/esporte/30437-custo-final-de-jogos-deve-estourar.shtml>
. Archived on:
<https://web.archive.org/web/20230918191156/https://www1.folha.uol.com.br/fsp/esporte/30
437-custo-final-de-jogos-deve-estourar.shtml>. Archive date: 18th September 2023.

Folha (2012b). Olympic Park is returned to London and will receive almost BRL 1 billion
renovation. Available on:
<https://www1.folha.uol.com.br/esporte/1191981-parque-olimpico-e-devolvido-a-londres-e-v
ai-receber-reforma-de-quase-r-1-bi.shtml>. Archived on:
<https://web.archive.org/web/20230918191223/https://www1.folha.uol.com.br/esporte/2012/
11/1191981-parque-olimpico-e-devolvido-a-londres-e-vai-receber-reforma-de-quase-r-1-bi.sh
tml>. Archive date: 18th September 2023.

Folha (2017). Government cuts off 87% of the Ministry of Sports budget to 2018'. Available
on:
<http://www1.folha.uol.com.br/esporte/2017/09/1919798-governo-corta-87-do-orcamento-do
-ministerio-do-esporte-para-2018.shtml>. Archived on:
<https://web.archive.org/web/20230918184155/https://www1.folha.uol.com.br/esporte/2017/
09/1919798-governo-corta-87-do-orcamento-do-ministerio-do-esporte-para-2018.shtml>.
Archive date: 18th September 2023.

Folha (2017b). With debt of R$ 130 million, Rio 2016 asks governments for help. Available
on:

<http://www1.folha.uol.com.br/esporte/2017/08/1907974-com-divida-de-r-130-milhoes-rio-2 016-pede-socorro-a-governos.shtml>. Archived on: <https://web.archive.org/web/20230918185509/https://www1.folha.uol.com.br/esporte/2017/ 08/1907974-com-divida-de-r-130-milhoes-rio-2016-pede-socorro-a-governos.shtml>. Archive date: 18th September 2023.

Folha (2020). Quatro anos depois, cariocas têm avaliação negativa da Olimpíada do Rio. Folha. Available on: <https://www1.folha.uol.com.br/esporte/2020/11/quatro-anos-depois-cariocas-tem-avaliacao-negativa-da-olimpiada-do-rio.shtml>. Archived on: <https://web.archive.org/web/20220228153029/https://www1.folha.uol.com.br/esporte/2020/ 11/quatro-anos-depois-cariocas-tem-avaliacao-negativa-da-olimpiada-do-rio.shtml>. Archive date: 15th January 2022.

Gaier, R. (2016). Justiça do Rio determina apresentação de plano de legado olímpico até 4 de agosto. UOL. Available on:<https://noticias.uol.com.br/ultimas-noticias/reuters/2016/07/20/justica-do-rio-determina-a presentacao-de-plano-de-legado-olimpico-ate-4-de-agosto.htm>. Archived on: <https://web.archive.org/web/20160721124528/https://noticias.uol.com.br/ultimas-noticias/re uters/2016/07/20/justica-do-rio-determina-apresentacao-de-plano-de-legado-olimpico-ate-4-d e-agosto.htm>. Archive date: 21th September 2016.

Gaier, R. (2017). New fire breaks out on the roof of the Rio Olympic Park Velodrome. Globo. Available on: <https://extra.globo.com/esporte/velodromo-olimpico-sofre-novo-incendio-no-rio-22115723. html>. Archived on: <https://web.archive.org/web/20230918181452/https://extra.globo.com/esporte/velodromo-ol impico-sofre-novo-incendio-no-rio-22115723.html>. Archive date: 18th September 2023.

Gibson, O. (2011). Lord Coe: Olympic stadium will not be white elephant after London 2012. The Guardian. Available on: <https://www.theguardian.com/sport/2011/oct/19/olympic-stadium-london-2012>. Archived on: <https://web.archive.org/web/20230918173337/https://www.theguardian.com/sport/2011/oct/ 19/olympic-stadium-london-2012>. Archive date: 18th September 2023.

Gibson, O. (2012). Boris Johnson takes charge of Olympic Park's future. The Guardian. Available on: <https://www.theguardian.com/politics/2012/sep/12/boris-johnson-olympic-park-legacy>. Archived on: <https://web.archive.org/web/20230918175059/https://www.theguardian.com/politics/2012/s ep/12/boris-johnson-olympic-park-legacy>. Archive date: 18th September 2023.

Gibson, O. (2012b). London 2012 Olympics will cost a total of £8.921bn, says minister. The Guardian. Available on: <https://www.theguardian.com/sport/2012/oct/23/london-2012-olympics-cost-total>.

Archived on:
<https://web.archive.org/web/20220228150840/https://www.theguardian.com/sport/2012/oct/
23/london-2012-olympics-cost-total>. Archive date: 11th January, 2022.

Gibson, O. (2013). London 2012 Olympic Games legacy questioned as young people shun
sport. The Guardian. Available on:
<https://www.theguardian.com/sport/2013/dec/12/london-2012-olympic-legacy-young-peopl
e-shun-sport>. Archived on:
<https://web.archive.org/web/20230918170101/https://www.theguardian.com/sport/2013/dec/
12/london-2012-olympic-legacy-young-people-shun-sport>. Archive date: 18th September
2023.

Gibson, O. (2015). Olympic legacy failure: inspiring London 2012 message has become a
millstone. The Guardian. Available on:
<https://www.theguardian.com/sport/blog/2015/jul/05/olympic-legacy-failure-london-2012-m
essage-millstone>. Archived on:
<https://web.archive.org/web/20230918174510/https://www.theguardian.com/sport/blog/201
5/jul/05/olympic-legacy-failure-london-2012-message-millstone>. Archive date: 18th
September 2023.

Globo (2011). Rio's Olympic budget has to be defined now, says British Minister. Available
on:
<http://g1.globo.com/brasil/noticia/2011/06/orcamento-olimpico-ja-tem-de-ser-definido-no-ri
o-diz-ministro-britanico.html>. Archived on:
<https://web.archive.org/web/20230918180048/https://g1.globo.com/brasil/noticia/2011/06/o
rcamento-olimpico-ja-tem-de-ser-definido-no-rio-diz-ministro-britanico.html>. Archive date:
18th September 2023.

Globo (2012). London 2012 will be the model for Olympic future, says Rogge. Available on:
<https://extra.globo.com/esporte/londres-2012-sera-modelo-para-futuro-olimpico-diz-rogge-4
438346.html>. Archived on:
<https://web.archive.org/web/20230918190834/https://extra.globo.com/esporte/londres-2012-
sera-modelo-para-futuro-olimpico-diz-rogge-4438346.html>. Archive date: 18th September
2023.

Globo (2012b). Olimpíada é 'desculpa fantástica' para mudar o Rio, diz prefeito. Available
on:
<https://g1.globo.com/brasil/noticia/2012/03/olimpiada-e-desculpa-fantastica-para-mudar-o-r
io-diz-prefeito.html>. Archived on:
<https://web.archive.org/web/20230918180156/https://g1.globo.com/brasil/noticia/2012/03/o
limpiada-e-desculpa-fantastica-para-mudar-o-rio-diz-prefeito.html>. Archive date: 18th
September 2023.

Globo (2012c). School named after Socrates will be part of Olympic legacy. Available on:
<https://globoesporte.globo.com/olimpiadas/noticia/2012/04/escola-com-nome-de-socrates-fa
ra-parte-do-projeto-de-legado-olimpico.html>. Archived on:

<https://web.archive.org/web/20230918180426/https://ge.globo.com/olimpiadas/noticia/2012/04/escola-com-nome-de-socrates-fara-parte-do-projeto-de-legado-olimpico.html>. Archive date: 18th September 2023.

Globo (2013). Cidades e Soluções follows works in Rio for the 2016 Olympics. Available on: <http://g1.globo.com/globo-news/noticia/2013/12/cidades-e-solucoes-acompanha-obras-para-olimpiadas-de-2016.html>. Archived on: <https://web.archive.org/web/20230918185216/https://g1.globo.com/globo-news/noticia/2013/12/cidades-e-solucoes-acompanha-obras-para-olimpiadas-de-2016.html>. Archive date: 18th September 2023.

Globo (2013b). Sanitation becomes an Olympic challenge because of the Guanabara Bay. Available on: <https://g1.globo.com/jornal-nacional/noticia/2013/10/saneamento-vira-desafio-olimpico-por-causa-da-baia-de-guanabara.html>. Archived on: <https://web.archive.org/web/20230918180819/https://g1.globo.com/jornal-nacional/noticia/2013/10/saneamento-vira-desafio-olimpico-por-causa-da-baia-de-guanabara.html>. Archive date: 18th September 2023.

Globo (2014). Lars espera legado olímpico próximo do zero e defende vela em Búzios. Available on: <https://globoesporte.globo.com/olimpiadas/noticia/2014/04/lars-espera-legado-olimpico-proximo-de-zero-e-defende-vela-em-buzios.html>. Archived on: <https://web.archive.org/web/20230918180717/https://ge.globo.com/olimpiadas/noticia/2014/04/lars-espera-legado-olimpico-proximo-de-zero-e-defende-vela-em-buzios.html>. Archive date: 18th September 2023.

Globo (2015). Eduardo Paes takes a stand against the impeachment of Dilma Rousseff. Available on: <https://g1.globo.com/globo-news/noticia/2015/12/eduardo-paes-se-posiciona-contra-o-impeachment-de-dilma-rousseff.html>. Archived on: <https://web.archive.org/web/20230918183752/https://g1.globo.com/globo-news/noticia/2015/12/eduardo-paes-se-posiciona-contra-o-impeachment-de-dilma-rousseff.html>. Archive date: 18th September 2023.

Globo (2016). At the end of Rio 2016, Government will keep investments in sports. Available on: <http://globoesporte.globo.com/olimpiadas/noticia/2016/08/no-fim-da-rio-2016-governo-garante-manter-investimentos-no-esporte.html>. Archived on: <https://web.archive.org/web/20230918184113/https://ge.globo.com/olimpiadas/noticia/2016/08/no-fim-da-rio-2016-governo-garante-manter-investimentos-no-esporte.html>. Archive date: 18th September 2023.

Globo (2016b). Olympic legacy: Rio community wins volleyball training centre. *Globo*. Available on: <http://ge.globo.com/olimpiadas/volei/noticia/2016/08/legado-olimpico-comunidade-do-rio-g

anha-centro-de-treinamento-de-volei.html>. Archived on:
<https://web.archive.org/web/20220228151528/http://ge.globo.com/olimpiadas/volei/noticia/
2016/08/legado-olimpico-comunidade-do-rio-ganha-centro-de-treinamento-de-volei.html>.
Archive date: 15th January, 2022.

Globo (2017a). Fire breaks out in Rio de Janeiro Velodrome. Available on:
<https://cbn.globoradio.globo.com/media/audio/107645/incendio-atinge-velodromo-do-rio-de
-janeiro.htm>. Archived on:
<https://web.archive.org/web/20230918181250/https://cbn.globoradio.globo.com/media/audi
o/107645/incendio-atinge-velodromo-do-rio-de-janeiro.htm>. Archive date: 18th September
2023.

Globo (2017b). One year after Rio 2016, Olympic arenas are underused. Available on:
<https://epocanegocios.globo.com/Brasil/noticia/2017/08/um-ano-depois-da-rio-2016-arenas-
olimpicas-sao-subutilizadas.html>. Archived on:
<https://web.archive.org/web/20230918185255/https://epocanegocios.globo.com/Brasil/notic
ia/2017/08/um-ano-depois-da-rio-2016-arenas-olimpicas-sao-subutilizadas.html>. Archive
date: 18th September 2023.

Globo (2017c). Preliminary investigation points to external cause of fire at Velodrome.
Available on:
<https://globoesporte.globo.com/olimpiadas/noticia/pericia-preliminar-aponta-causa-externa-
em-incendio-no-velodromo.ghtml>. Archived
on:<https://web.archive.org/web/20230918181416/https://ge.globo.com/olimpiadas/noticia/p
ericia-preliminar-aponta-causa-externa-em-incendio-no-velodromo.ghtml>. Archive date:
18th September 2023.

Globo (2019). The Essential Morning: the melancholy inheritance of the Olympics. O Globo.
Available on:
<https://oglobo.globo.com/politica/o-essencial-da-manha-melancolica-heranca-das-olimpiada
s-24086274>. Archived on:
<https://web.archive.org/web/20220228145601/https://oglobo.globo.com/politica/o-essencial
-da-manha-melancolica-heranca-das-olimpiadas-24086274>. Archive date: 10th January,
2022.

Globo (2019b). Decree creates office to manage the Olympic legacy of the Rio Games in
2016. Globo Esporte. Available on:
<https://globoesporte.globo.com/olimpiadas/noticia/decreto-cria-escritorio-para-gestao-de-le
gado-olimpico-dos-jogos-do-rio-em-2016.ghtml>. Archived on:
<https://web.archive.org/web/20231014165250/https://ge.globo.com/olimpiadas/noticia/decre
to-cria-escritorio-para-gestao-de-legado-olimpico-dos-jogos-do-rio-em-2016.ghtml>. Archive
date: 14th October 2023.

Globo (2020). PagSeguro's profit risis 50.2% in 2019. Available on:
<https://valor.globo.com/financas/noticia/2020/02/27/lucro-da-pagseguro-avanca-502percent-
em-2019.ghtml>. Archived on:

<https://web.archive.org/web/20230918185422/https://valor.globo.com/financas/noticia/2020/02/27/lucro-da-pagseguro-avanca-502percent-em-2019.ghtml>. Archive date: 18th September 2023.

Globo (2021). Olympic legacy: Rio 2016 Tennis Center goes to the private sector. Available on: <https://ge.globo.com/tenis/noticia/legado-olimpico-centro-de-tenis-da-rio-2016-vai-para-a-iniciativa-privada.ghtml>. Archived on: <https://web.archive.org/web/20230918181339/https://ge.globo.com/tenis/noticia/legado-olimpico-centro-de-tenis-da-rio-2016-vai-para-a-iniciativa-privada.ghtml>. Archive date: 18th September 2023.

Globo (2021b). Olympic legacy: Rio 2016 Tennis Centre goes to the private sector. Available on: <https://ge.globo.com/tenis/noticia/legado-olimpico-centro-de-tenis-da-rio-2016-vai-para-a-iniciativa-privada.ghtml>. Archived on: <https://web.archive.org/web/20230918181339/https://ge.globo.com/tenis/noticia/legado-olimpico-centro-de-tenis-da-rio-2016-vai-para-a-iniciativa-privada.ghtml>. Archive date: 18th September 2023.

Gozzer, T. (2019). With extinction in sight, AGLO projects Rio 2016's Olympic legacy in check. Globo. Available on: <https://ge.globo.com/olimpiadas/noticia/para-presidente-extincao-da-aglo-representaria-o-pior-no-esporte-seria-o-fim-do-legado-olimpico.ghtml>. Archived on: <https://web.archive.org/web/20230918183934/https://ge.globo.com/olimpiadas/noticia/para-presidente-extincao-da-aglo-representaria-o-pior-no-esporte-seria-o-fim-do-legado-olimpico.ghtml>. Archive date: 18th September 2023.

Grandjean, G. (2012). London 2012 legacy, The Olympics has cost me my home. The Guardian. Available on: <https://www.theguardian.com/sport/video/2012/jun/13/london-2012-legacy-olympics-homes-video>. Archived on: <https://web.archive.org/web/20230918175738/https://www.theguardian.com/sport/video/2012/jun/13/london-2012-legacy-olympics-homes-video>. Archive date: 18th September 2023.

Guerra, M. (2015). 14 months to Rio 2016, uncertainty over main legacy's future. Globo. Available on: <https://globoesporte.globo.com/olimpiadas/noticia/2015/05/14-meses-par-ao-rio-2016-principal-legado-esportivo-tem-futuro-incerto.html>. Archived on: <https://web.archive.org/web/20230918180354/https://ge.globo.com/olimpiadas/noticia/2015/05/14-meses-par-ao-rio-2016-principal-legado-esportivo-tem-futuro-incerto.html>. Archive date: 18th September 2023.

Hart, S. (2010). Government accused of 'devastating' legacy with school spending cuts. Telegraph. Available on: <https://www.telegraph.co.uk/sport/othersports/schoolsports/8076966/London-2012-Olympic

s-Government-accused-of-devastating-legacy-with-school-spending-cuts.html>. Archived on: <https://web.archive.org/web/20230918174217/https://www.telegraph.co.uk/sport/othersports /schoolsports/8076966/London-2012-Olympics-Government-accused-of-devastating-legacy-with-school-spending-cuts.html>. Archive date: 18th September 2023.

Hill, D. (2010). Uncertainty about Olympic legacy continues. The Guardian. Available on: <https://www.theguardian.com/uk/davehillblog/2010/mar/08/londn-2012-olympics-kate-hoey -legacy>. Archived on: <https://web.archive.org/web/20230918171404/https://www.theguardian.com/uk/davehillblo g/2010/mar/08/londn-2012-olympics-kate-hoey-legacy>. Archive date: 18th September 2023.

Hill, D. (2012). Stratford's Olympic legacy: It's been far better than I thought. The Guardian. Available on: <https://www.theguardian.com/uk/2012/aug/09/stratford-olympic-legacy-better>. Archived on: <https://web.archive.org/web/20230918175859/https://www.theguardian.com/uk/2012/aug/0 9/stratford-olympic-legacy-better>. Archive date: 18th September 2023.

Hone , D. (2012). The Olympic legacy will be a festival of the people. The Guardian. Available on: <https://www.theguardian.com/commentisfree/2012/feb/06/olypic-legacy-festival-of-people> . Archived on: <https://web.archive.org/web/20230918175930/https://www.theguardian.com/commentisfree/ 2012/feb/06/olypic-legacy-festival-of-people>. Archive date: 18th September 2023.

Hunt, E. (2016). Why is the Olympic diving pool green? The good news is it's not urine. The Guardian. Available on: <https://www.theguardian.com/sport/2016/aug/10/why-has-the-olympic-pool-turned-green-th e-good-news-is-its-not-urine>. Archived on: <https://web.archive.org/web/20230918181056/https://www.theguardian.com/sport/2016/aug /10/why-has-the-olympic-pool-turned-green-the-good-news-is-its-not-urine>. Archive date: 18th September 2023.

Intrator, S. (2016). Olympic heritage for cariocas. Globo. Available on: <https://epoca.globo.com/esporte/olimpiadas/noticia/2016/06/heranca-olimpica-para-os-cario cas.html>. Archived on: <https://web.archive.org/web/20230918185921/https://epoca.globo.com/esporte/olimpiadas/n oticia/2016/06/heranca-olimpica-para-os-cariocas.html>. Archive date: 18th September 2023.

Kelso, P. (2009). London 2012: Serious questions remain about Olympic legacy. Telegraph. Available on: <https://www.telegraph.co.uk/sport/olympics/london-2012/5758718/London-2012-Serious-q uestions-remain-about-Olympic-legacy.html>. Archived on: <https://web.archive.org/web/20230918171300/https://www.telegraph.co.uk/sport/olympics/l ondon-2012/5758718/London-2012-Serious-questions-remain-about-Olympic-legacy.html>. Archive date: 18th September 2023.

Lemos , R. (2011). Meirelles diz que Rio 2016 não repetirá erros do Pan 2007.Veja. Available on: <https://veja.abril.com.br/esporte/meirelles-diz-que-rio-2016-nao-repetira-erros-do-pan-2007>. Archived on: <https://web.archive.org/web/20230918180115/https://veja.abril.com.br/esporte/meirelles-diz-que-rio-2016-nao-repetira-erros-do-pan-2007>. Archive date: 18th September 2023.

London, B. (2014). Prince Harry and Boris Johnson enjoy a playdate at the Olympic Park. Daily Mail. Available on: <https://www.dailymail.co.uk/femail/article-2596797/Still-reeling-zip-wire-incident-Boris-Mayor-takes-backseat-Prince-Harry-pay-visit-new-Olympic-Park.html>. Archived on: <https://web.archive.org/web/20230918175529/https://www.dailymail.co.uk/femail/article-2596797/Still-reeling-zip-wire-incident-Boris-Mayor-takes-backseat-Prince-Harry-pay-visit-new-Olympic-Park.html>. Archive date: 18th September 2023.

Lusher, A. (2008). Olympics: youth sports struggle as British win gold at Beijing Games. Telegraph. Available on: <https://www.telegraph.co.uk/sport/olympics/london-2012/2609855/Olympics-Youth-sports-struggle-as-British-win-gold-at-Beijing-Games.html>. Archived on: <https://web.archive.org/web/20230918174037/https://www.telegraph.co.uk/sport/olympics/london-2012/2609855/Olympics-Youth-sports-struggle-as-British-win-gold-at-Beijing-Games.html>. Archive date: 18th September 2023.

McCurry, J. (2020). A year to go, but does Tokyo still want to host the Olympics?. The Guardian. Available on: <https://www.theguardian.com/world/2020/jul/19/a-year-to-go-but-does-tokyo-still-want-to-host-the-olympics-after-covid?utm_source=dlvr.it&utm_medium=twitter>. Archived on: <https://web.archive.org/web/20231014170350/https://www.theguardian.com/world/2020/jul/19/a-year-to-go-but-does-tokyo-still-want-to-host-the-olympics-after-covid>. Archive date: 14th October 2023.

Meredith, D. (2012). The Olympic legacy: creating a new community for London in Stratford. The Guardian. Available on: <https://www.theguardian.com/housing-network/2012/jul/19/olympic-legacy-social-housing-east-village>. Archived on: <https://web.archive.org/web/20230918175840/https://www.theguardian.com/housing-network/2012/jul/19/olympic-legacy-social-housing-east-village>. Archive date: 18th September 2023.

Milliken and Haynes (2016). On tight budget, opening ceremony adopts 'MacGyver' approach. Reuters. Available on: <https://www.reuters.com/article/uk-olympics-rio-opening-idUKKCN10F2FK>. Archived on: <https://web.archive.org/web/20231014204439/https://www.reuters.com/article/uk-olympics-rio-opening-idUKKCN10F2FK>. Archive date 14th October 2023.

Nitahara, A. (2021). Dies, aged 66, Jorge Picciani, former president of the Assembly of Rio. Agencia Brasil. Available on: <https://agenciabrasil.ebc.com.br/politica/noticia/2021-05/morre-aos-66-anos-jorge-picciani-ex-presidente-da-assembleia-do-rio>. Archived on: <https://web.archive.org/web/20230918182834/https://agenciabrasil.ebc.com.br/politica/notic ia/2021-05/morre-aos-66-anos-jorge-picciani-ex-presidente-da-assembleia-do-rio>. Archive date: 18th September 2023.

Nogueira, I. (2016). Outside the election, Paes runs to church', says Crivella. Folha. Available on: <https://www1.folha.uol.com.br/poder/eleicoes-2016/2016/09/1817909-fora-da-eleicao-paes-corre-para-igreja-diz-crivella.shtml>. Archived on: <https://web.archive.org/web/20230918182532/https://www1.folha.uol.com.br/poder/eleicoe s-2016/2016/09/1817909-fora-da-eleicao-paes-corre-para-igreja-diz-crivella.shtml>. Archive date: 18th September 2023.

O'Carroll, L. (2019). Will getting Brexit done create another baby boom in the UK?. The Guardian. Available on: <https://www.theguardian.com/politics/2019/dec/08/will-getting-brexit-done-create-another-b aby-boom-in-the-uk-factcheck>. Archived on: <https://web.archive.org/web/20230918174844/https://www.theguardian.com/politics/2019/d ec/08/will-getting-brexit-done-create-another-baby-boom-in-the-uk-factcheck>. Archive date: 18th September 2023.

Ofcom (2022). News Consumption in the UK: 2022. Available on: <https://www.ofcom.org.uk/__data/assets/pdf_file/0027/241947/News-Consumption-in-the-UK-2022-report.pdf>. Archived on: <https://web.archive.org/web/20230720110320/https://www.ofcom.org.uk/__data/assets/pdf_ file/0027/241947/News-Consumption-in-the-UK-2022-report.pdf>. Archive date: 20 July 2023.

Olympic Channel (2016). Opening Ceremony - Rio 2016 Olympic Games. Olympic Broadcasting Services. Available on: <https://olympics.com/en/video/opening-ceremony-rio-2016-olympic-games>. Accessed on 14th October 2023. [VIDEO]

Parry, H. (2016). Rio Olympics could be under threat as Brazil riots continue over political unrest. Daily Mail. Available on: <https://www.dailymail.co.uk/news/article-3588239/Fears-Olympics-Brazil-plunges-political -chaos-Riot-police-clash-activists-streets-acting-president-installs-SECOND-sports-minister-t wo-months.html>. Archived on: <https://web.archive.org/web/20230918190630/https://www.dailymail.co.uk/news/article-358 8239/Fears-Olympics-Brazil-plunges-political-chaos-Riot-police-clash-activists-streets-acting -president-installs-SECOND-sports-minister-two-months.html>. Archive date: 18th September 2023.

Pereira, P. (2014). Olympic dream changes Rio's landscape. Estadao. Available on: <https://esportes.estadao.com.br/noticias/jogos-olimpicos,sonho-do-ouro-olimpico-muda-paisagem-do-rio,1606378>. Archived on: <https://web.archive.org/web/20230918180523/https://www.estadao.com.br/esportes/olimpiadas/sonho-do-ouro-olimpico-muda-paisagem-do-rio/>. Archive date: 18th September 2023.

Pimm, M. (2012). Letters: The true Olympic legacy is white elephants on our doorstep. The Guardian. Available on: <https://www.theguardian.com/sport/2012/dec/27/true-olympic-legacy-white-elephants.>. Archived on: <https://web.archive.org/web/20230918195949/https://www.theguardian.com/sport/2012/dec/27/true-olympic-legacy-white-elephants>. Archive date: 18th September 2023.

Press Association (2012). Olympic legacy: school sports provision patchy across UK, admits Jeremy Hunt. The Guardian. Available on: <https://www.theguardian.com/sport/2012/aug/06/olympic-legacy-school-sports-provision>. Archived on: <https://web.archive.org/web/20220224144407/https://www.theguardian.com/sport/2012/aug/06/olympic-legacy-school-sports-provision>. Accessed on 14th January, 2022.

Pussieldi, A. (2018). Six years later, the wonderful legacy of London 2012. Globo. Available on: <https://sportv.globo.com/site/blogs/blog-do-coach/post/2018/08/01/seis-anos-depois-o-belo-legado-olimpico-de-londres-2012.ghtml>. Archived on: <https://web.archive.org/web/20230918190959/https://sportv.globo.com/site/blogs/blog-do-coach/post/2018/08/01/seis-anos-depois-o-belo-legado-olimpico-de-londres-2012.ghtml>. Archive date: 18th September 2023.

Rede do Esporte (2015). Cultural and mobility legacies will be hallmarks of 2016, say experts. Available on: <http://rededoesporte.gov.br/pt-br/noticias/legados-culturais-e-de-mobilidade-serao-marcas-dos-jogos-de-2016-preveem-especialistas>. Archived on: <https://web.archive.org/web/20230918182703/http://rededoesporte.gov.br/pt-br/noticias/legados-culturais-e-de-mobilidade-serao-marcas-dos-jogos-de-2016-preveem-especialistas>. Archive date: 18th September 2023.

Rede do Esporte (2017). Brazil shines in the opening of the stage of the world circuit held at the Olympic Park. Available on: <http://rededoesporte.gov.br/pt-br/noticias/com-jogos-olimpicos-na-memoria-brasil-brilha-no-primeiro-dia-da-etapa-do-rj-do-circuito-mundial-de-volei-de-praia>. Archived on: <https://web.archive.org/web/20230918183843/http://rededoesporte.gov.br/pt-br/noticias/com-jogos-olimpicos-na-memoria-brasil-brilha-no-primeiro-dia-da-etapa-do-rj-do-circuito-mundial-de-volei-de-praia>. Archive date: 18th September 2023.

Rede do Esporte (2017b). Event at Barra Olympic Park celebrates one year of the Games and the legacy of the event to the country. Available on:

<http://rededoesporte.gov.br/pt-br/noticias/evento-no-parque-olimpico-da-barra-comemora-um-ano-dos-jogos-e-o-legado-do-evento-ao-pais>. Archived on: <https://web.archive.org/web/20230918183140/http://rededoesporte.gov.br/pt-br/noticias/evento-no-parque-olimpico-da-barra-comemora-um-ano-dos-jogos-e-o-legado-do-evento-ao-pais>. Archive date: 18th September 2023.

Rede do Esporte (2017c). Ricardo Leyser: The Games are not the point of arrival, but the point of departure. Available on: <http://rededoesporte.gov.br/pt-br/noticias/ricardo-leyser-201colimpiadas-nao-sao-ponto-de-chegada-mas-de-partida201d>. Archived on: <https://web.archive.org/web/20230918184042/http://rededoesporte.gov.br/pt-br/noticias/ricardo-leyser-201colimpiadas-nao-sao-ponto-de-chegada-mas-de-partida201d>. Archive date: 18th September 2023.

Reporters without Borders (2023). Brazil. Available on: <https://rsf.org/en/country/brazil>. Archived on: <https://web.archive.org/web/20230916142558/https://rsf.org/en/country/brazil>. Archive date: 16th September 2023.

Reuters Institute (2022). Digital News Report. Available on: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf>. Archived on: <https://web.archive.org/web/20231011124214/https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf>. Archive date: 11th January 2023.

Ronay, B. (2022). Legacy of London 2012 is one that has sold Britain's fitness short. The Guardian. Available on: <https://www.theguardian.com/sport/blog/2022/jun/18/legacy-of-london-2012-is-one-that-has-sold-britains-fitness-short>. Archived on: <https://web.archive.org/web/20230918174821/https://www.theguardian.com/sport/blog/2022/jun/18/legacy-of-london-2012-is-one-that-has-sold-britains-fitness-short>. Archive date: 18th September 2023.

Rumsby, B. (2022). UK Athletics in talks to quit London Stadium and open door for West Ham transformation. Telegraph. Available on: <https://www.telegraph.co.uk/football/2022/05/05/uk-athletics-talks-quit-london-stadium-open-door-west-ham-transformation/>. Archived on: <https://web.archive.org/web/20230918173536/https://www.telegraph.co.uk/football/2022/05/05/uk-athletics-talks-quit-london-stadium-open-door-west-ham-transformation/>. Archive date: 18th September 2023.

Rumsby, B. (2022b). Were the London 2012 Olympics a failure?. Telegraph. Available on: <https://www.telegraph.co.uk/olympics/2022/07/27/london-2012-olympics-failure/>. Archived on: <https://web.archive.org/web/20230918174350/https://www.telegraph.co.uk/olympics/2022/07/27/london-2012-olympics-failure/>. Archive date: 18th September 2023.

Samuel, M. (2010). An Olympic medal in utter madness as Tessa Jowell aims to deliver Hammer blow to West Ham's future Stratford move. Daily Mail. Available on: <https://www.dailymail.co.uk/sport/article-1252770/Martin-Samuel-Tessa-Jowell-deserves-Olympic-medal-utter-madness.html>. Archived on: <https://web.archive.org/web/20230918173558/https://www.dailymail.co.uk/sport/article-1252770/Martin-Samuel-Tessa-Jowell-deserves-Olympic-medal-utter-madness.html>. Archive date: 18th September 2023.

Samuel, M. (2013). Martin Samuel: West Ham saved the Olympic Stadium from the idiots who built it, no wonder Karren Brady is smiling. Daily Mail. Available on: <https://www.dailymail.co.uk/sport/football/article-2299930/MARTIN-SAMUEL--THE-DEBATE-West-Ham-saved-Olympic-Stadium-idiots-built-wonder-Karren-Brady-smiling.html>. Archived on: <https://web.archive.org/web/20230918175131/https://www.dailymail.co.uk/sport/football/article-2299930/MARTIN-SAMUEL--THE-DEBATE-West-Ham-saved-Olympic-Stadium-idiots-built-wonder-Karren-Brady-smiling.html>. Archive date: 18th September 2023.

Scott, M. (2008). London 2012 Olympic stadium may not have an athletics legacy after all. The Guardian. Available on: <https://www.theguardian.com/uk/2008/aug/21/olympics2012.olympics2008>. Archived on: <https://web.archive.org/web/20230918170247/https://www.theguardian.com/uk/2008/aug/21/olympics2012.olympics2008>. Archive date: 18th September 2023.

Soares, L. (2016). Environmental legacy: the great debt of Rio Olympics. Globo. Available on: <https://cbn.globoradio.globo.com/grandescoberturas/rio-2016/2016/07/02/LEGADO-AMBIENTAL-A-GRANDE-DIVIDA-DA-OLIMPIADA-DO-RIO.htm>. Archived on: <https://web.archive.org/web/20230918181133/https://cbn.globoradio.globo.com/grandescoberturas/rio-2016/2016/07/02/LEGADO-AMBIENTAL-A-GRANDE-DIVIDA-DA-OLIMPIADA-DO-RIO.htm>. Archive date: 18th September 2023.

Sousa, A. (2013). Herança bendita. Globo. Available on: <https://oglobo.globo.com/cultura/heranca-bendita-9490662>. Archived on: <https://web.archive.org/web/20230918191304/https://oglobo.globo.com/cultura/heranca-bendita-9490662>. Archive date: 18th September 2023.

Smith, M. (2017). How left or right-wing are the UK's newspapers? Available on: <https://yougov.co.uk/topics/politics/articles-reports/2017/03/07/how-left-or-right-wing-are-uks-newspapers>. Archived on: <https://web.archive.org/web20230709161602/https://yougov.co.uk/topics/politics/articles-reports/2017/03/07/how-left-or-right-wing-are-uks-newspapers>. Archive date: 9th September 2023.

Telegraph (2011). Olympic Stadium will become 'desolate graveyard if West Ham win battle with Tottenham, says Harry Redknapp. Available on: <https://www.telegraph.co.uk/sport/olympics/8273298/Olympic-Stadium-will-become-desola

te-graveyard-if-West-Ham-win-battle-with-Tottenham-says-Harry-Redknapp.html>. Archived on: <https://web.archive.org/web/20230918173940/https://www.telegraph.co.uk/sport/olympics/8 273298/Olympic-Stadium-will-become-desolate-graveyard-if-West-Ham-win-battle-with-Tot tenham-says-Harry-Redknapp.html>. Archive date: 18th September 2023.

Telegraph (2007). Olympics 2012: what will happen when the crowds leave. Available on: <https://www.telegraph.co.uk/finance/personalfinance/2817878/Olympics-2012-what-will-ha ppen-when-the-crowds-leave.html>. Archived on: <https://web.archive.org/web/20230918174635/https://www.telegraph.co.uk/finance/personal finance/2817878/Olympics-2012-what-will-happen-when-the-crowds-leave.html>. Archive date: 18th September 2023.

Telegraph (2013). Boris Johnson hails 'momentous' West Ham Olympic Stadium deal. Available on: <https://www.telegraph.co.uk/sport/football/teams/west-ham/9948379/Boris-Johnson-hails-m omentous-West-Ham-Olympic-Stadium-deal.html>. Archived on: <https://web.archive.org/web/20130425004257/https://www.telegraph.co.uk/sport/football/te ams/west-ham/9948379/Boris-Johnson-hails-momentous-West-Ham-Olympic-Stadium-deal. html>. Archive date: 25th April 2013.

Telegraph (2019). Boris Johnson is Mr Brexit. Elect him PM and give him a chance to deliver it. Available on: <https://www.telegraph.co.uk/opinion/2019/07/04/boris-johnson-mr-brexit-elect-pm-give-cha nce-deliver/>. Archived on: <https://web.archive.org/web/20190815172440/https://www.telegraph.co.uk/opinion/2019/07 /04/boris-johnson-mr-brexit-elect-pm-give-chance-deliver/>. Archive date: 15th August 2023.

The Guardian (2005). London Games would 'inspire athletes of tomorrow'. Available on: <https://www.theguardian.com/uk/2005/jul/06/olympics2012.olympicgames>. Archived on: <https://web.archive.org/web/20230918174008/https://www.theguardian.com/uk/2005/jul/06/ olympics2012.olympicgames>. Archive date: 18th September 2023.

The Guardian (2012). Olympic legacy: school sport provision patchy across UK, admits Jeremy Hunt. Available on: <https://www.theguardian.com/sport/2012/aug/06/olympic-legacy-school-sports-provision>. Archived on: <https://web.archive.org/web/20230918174445/https://www.theguardian.com/sport/2012/aug /06/olympic-legacy-school-sports-provision>. Archive date: 18th September 2023.

The Guardian (2016). Who owns the Guardian, and how is it funded? Available on: <https://www.theguardian.com/membership/video/2016/nov/11/who-owns-the-guardian-and-how-is-it-funded>. Archived on: <https://web.archive.org/web/20221114232656/https://www.theguardian.com/membership/vi

deo/2016/nov/11/who-owns-the-guardian-and-how-is-it-funded>. Achieve date: 14th November 2022.

The Guardian (2017). They smashed my face and demolished my home: all for the Rio Olympics. Available on: <https://www.theguardian.com/global-development/2017/may/18/smashed-my-face-demolished-my-home-rio-olympic-games-maria-da-penha-macena>. Archived on: <https://web.archive.org/web/20230918185039/https://www.theguardian.com/global-development/2017/may/18/smashed-my-face-demolished-my-home-rio-olympic-games-maria-da-penha-macena>. Archive date: 18th September 2023.

Uchoa, P. (2016). Maria da Penha: The woman who changed Brazil's domestic violence laws. BBC. Available on: <https://www.bbc.co.uk/news/magazine-37429051>. Archived on: <https://web.archive.org/web/20230918185103/https://www.bbc.com/news/magazine-37429051>. Archive date: 18th September 2023.

UK Sport (2014). Coe's London 2012 International Inspiration legacy promise reaches 25 million around the world. Available on: <https://www.uksport.gov.uk/news/2014/07/01/coes-london-2012-international-inspiration-legacy-promise-reaches-25-million-around-the-world>. Archived on: <https://web.archive.org/web/20230918174421/https://www.uksport.gov.uk/news/2014/07/01/coes-london-2012-international-inspiration-legacy-promise-reaches-25-million-around-the-world>. Archive date: 18th September 2023.

UK Sport (2018). London 2012 legacy continues to provide a positive economic impact six years on. Available on: <https://www.uksport.gov.uk/news/2018/07/27/london-2012-legacy-continues-to-provide-a-positive-economic-impact-six-years-on>. Archive on: <https://web.archive.org/web/20231014164659/https://www.uksport.gov.uk/news/2018/07/27/london-2012-legacy-continues-to-provide-a-positive-economic-impact-six-years-on>. Archive date: 14th October 2023.

Varley, C. (2018). These haunting photos of abandoned Olympic sites tell their own stories. BBC. Available on: <https://www.bbc.co.uk/bbcthree/article/de4004d3-97e8-467c-89a9-03290074e34a>. Archived on: <https://web.archive.org/web/20230918190224/https://www.bbc.co.uk/bbcthree/article/de4004d3-97e8-467c-89a9-03290074e34a>. Archive date: 18th September 2023.

Wade, S. (2017). Rio 2016's venues abandoned and vandalised. Daily Mail. Available on: <https://www.dailymail.co.uk/sport/othersports/article-4210996/Rio-2016-s-venues-abandoned-vandalised.html>. Archived on: <https://web.archive.org/web/20230918190304/https://www.dailymail.co.uk/sport/othersports/article-4210996/Rio-2016-s-venues-abandoned-vandalised.html>. Archive date: 18th September 2023.

Warner, A. (2010). Mayor aims for total control of 2012 legacy. BBC. Available on: <https://www.bbc.co.uk/blogs/adrianwarner/2010/06/mayor_stakes_a_claim_for_contr.html>. Archived on: <https://web.archive.org/web/20230918173508/https://www.bbc.co.uk/blogs/adrianwarner/2010/06/mayor_stakes_a_claim_for_contr.html>. Archive date: 18th September 2023.

Warner, A. (2010b). Two Olympic sagas: Denials and broken promises. BBC. Available on: <https://www.bbc.co.uk/blogs/adrianwarner/2010/10/i_never_think_you_should.html>. Archived on: <https://web.archive.org/web/20230918173705/https://www.bbc.co.uk/blogs/adrianwarner/2010/10/i_never_think_you_should.html>. Archive date: 18th September 2023.

Watts, J. (2016). Zika crisis and economic woes bring gloom to Brazil's Olympic buildup. The Guardian . Available on: <https://www.theguardian.com/world/2016/feb/05/brazil-faces-steep-hurdles-ahead-of-olympics-zika-and-economic-crisis>. Archived on: <https://web.archive.org/web/20230918190717/https://www.theguardian.com/world/2016/feb/05/brazil-faces-steep-hurdles-ahead-of-olympics-zika-and-economic-crisis>. Archive date: 18th September 2023.

Wikipedia (2023). CPLP Games. Available on <https://en.wikipedia.org/wiki/CPLP_Games#:~:text=The%20CPLP%20Games%20(Portuguese%3A%20Jogos,less%20than%2016%20years%20old>. Archive on <https://web.archive.org/web/20230307191831/https://en.wikipedia.org/wiki/CPLP_Games#:~:text=The%20CPLP%20Games%20(Portuguese%3A%20Jogos,less%20than%2016%20years%20old>. Archive date: 7th March 2023.

Wright, J. (2021). Going for GOLD! How east London homes have shot up in value since 2012 Olympics with price of properties in Waltham Forest more than DOUBLING. Daily Mail. Available on: <https://www.dailymail.co.uk/news/article-9855507/How-east-London-homes-shot-value-2012-Olympics.html>. Archived on: <https://web.archive.org/web/20230918175637/https://www.dailymail.co.uk/news/article-9855507/How-east-London-homes-shot-value-2012-Olympics.html>. Archive date: 18th September 2023.

Yahoo Brasil (2016). Rio Justice determines presentation of Olympic legacy plan by August 4. Available on: <https://esportes.yahoo.com/noticias/justi%C3%A7a-rio-determina-apresenta%C3%A7%C3%A3o-plano-legado-ol%C3%ADmpico-at%C3%A9-192505368--spt.html?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAJk71D-prvJbiwa3HQvGVv5zTGFXlNof7-y7WFId8MCB0VxrpEGB2e7VPESawRRTfzSzVMfv1hBPjQufDY-uyBIvbsS7_MOeKUdMyJqUOd_HeSxuBwxM6VHhl2oBv8ilF_W201ic9H-ZEEHc9pjcdDwxvkbr7R5uhJw29dLNh0VA&guccounter=2>. Archived on:

<https://web.archive.org/web/20230918185820/https://br.search.yahoo.com/?fr2=p:sportsrd, mkt:br>. Archive date: 18th September 2023.

Secondary sources

Adami, M. & Kahn, G. (2023) What we learnt at our 40-year Fellowship reunion about the future of news. Available on <https://reutersinstitute.politics.ox.ac.uk/news/what-we-learnt-our-40-year-fellowship-reunion-about-future-news>. Accessed on 10th September 2023. Archived on <https://web.archive.org/web/20230911161427/https://reutersinstitute.politics.ox.ac.uk/news/what-we-learnt-our-40-year-fellowship-reunion-about-future-news>.

Alves, A. L. F., Baptista, C. de S., Firmino, A. A., de Oliveira, M. G., & de Figueirêdo, H. F. (2014). Temporal Analysis of Sentiment in Tweets: A Case Study with FIFA Confederations Cup in Brazil. In H. Decker, L. Lhotská, S. Link, M. Spies, & R. R. Wagner (Eds.), *Database and Expert Systems Applications* (Vol. 8644, pp. 81–88). Springer International Publishing. https://doi.org/10.1007/978-3-319-10073-9_7.

Al-Moslmi, T., Omar, N., Abdullah, S., & Albared, M. (2017). Approaches to cross-domain sentiment analysis: A systematic literature review. *Ieee access*, *5*, 16173-16192.

Aplin, N., & Lyon, L. M. W. (2014). The Youth Olympic Games in Singapore 2010: origins, impacts and legacy. In *The Youth Olympic Games* (pp. 199-215). Routledge.

Araújo M, Reis J, Pereira A, Benevenuto F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. In: *Proceedings of 31st ACM symposium on applied computing* (SAC).

Araújo M, Pereira A, Benevenuto F. (2020). A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Information Sciences*. https://doi.org/10.1016/j.ins.2019.10.031.

Austin, J. L. (1975). *How to do things with words*. Oxford University Press.

Bairner, A., & Molnar, G. (Eds.). (2010). The politics of the Olympics: A survey. Routledge.

Balahur, A., & Steinberger, R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *Proceeding of WOMSA*, *9*.

Balahur, A., & Turchi, M. (2012). Multilingual sentiment analysis using machine translation. In *Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis* (pp. 52-60).

Balahur, A. et al (2013). Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*.

Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 127-135).

Barthes, R., & Duisit, L. (1975). An introduction to the structural analysis of narrative. *New literary history*, *6*(2), 237-272.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In Third international AAAI conference on weblogs and social media.

Berger, P. L., & Luckmann, T. (1991). The social construction of reality: A treatise in the sociology of knowledge (No. 10). Penguin UK.

Berry, D. M., & Fagerjord, A. (2017). *Digital humanities: knowledge and critique in a digital age*. John Wiley & Sons.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. Reilly Media, Inc.

Black, M. L. (2016). The World Wide Web as complex data set: Expanding the Digital Humanities into the twentieth century and beyond through Internet research. *International Journal of Humanities and Arts Computing*, *10*(1), 95-109.

Borges, V. D. P. C., & Santos, N. (2018). Preparing tourist spaces for the 2014 Fifa World Cup thoughts about deterritorialisation, reterritorialisation and gentrification. Cuadernos de Turismo, (42), 93-117.

Boudana, S. (2016). Impartiality Is Not Fair: Toward an Alternative Approach to the Evaluation of Content Bias in News Stories. In *Journalism*, *17* (5): 600–618.

Bounegru, L. (n.d.). *Narrating Networks*. 33.

Bourdieu, P. (1992). Language and Symbolic power. Polity. 312p.

Bourdieu, P. (1998). On Television. New York: New York Press

Brady, W. et al. (2017). 'Emotion shapes the diffusion of moralized content in social networks'. *Proceedings of the National Academy of Sciences,* 114, 7313 - 7318.

Brank, J., Leban, G., Grobelnik, M (2017). Annotating Documents with Relevant Wikipedia Concepts. Proceedings of the Slovenian Conference on Data Mining and Data Warehouses (SiKDD 2017), Ljubljana, Slovenia.

Brügger, N., & Ditte, L. (2019). *The Historical Web and Digital Humanities: The case of national web domains*. Routledge.

Brügger, N. & Miiligan, I. (2019). The SAGE handbook of web history. London: SAGE publications.

Brügger, N., & Milligan, I. (2019). *The Web as History*. London: UCL Press.

Brügger, N. (2018). *The archived web: Doing history in the digital age*. MIT Press.

Brügger, N. (2017). Web 25: histories from the first 25 years of the World Wide Web. New York: Peter Lang Publishing.

Cairo, A. (2019). *How Charts Lie: Getting Smarter about Visual Information*. WW Norton & Company.

Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.

Cambria, E., Hussain, A. (2015) Sentic Computing: *A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. ISBN: 978-3-319-23654-4, Springer.

Chappelet, J. L. (2003). The legacy of the Olympic winter games: An overview. In *The legacy of the Olympic Games: 1984-2000: International Symposium, Lausanne, 14th, 15th and 16th November 2002* (pp. 54-66). International Olympic Committee.

Castro, D. G., & Novaes, P. R. (2014). Copa do Mundo 2014 e os impactos no direito à moradia: uma análise das cidades-sede brasileiras. OS IMPACTOS DA COPA DO MUNDO 2014 E DAS OLIMPÍADAS 2016, 20, 79.

Cioccari, D. (2015). Operação Lava Jato: escândalo, agendamento e enquadramento. *Revista Alterjor*, *12*(2), 58-78.

Cirqueira, D., Jacob, A., Lobato, F., Santana, A. L. D., & Pinheiro, M. (2016). Performance evaluation of sentiment analysis methods for Brazilian Portuguese. In International Conference on Business Information Systems (pp. 245-251). Springer, Cham.

Couldry, N., Hepp, A., & Krotz, F. (Eds.). (2010). *Media events in a global age*. Routledge.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., & Zhou, Q. (2016). Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. Cognitive computation, 8, 757–771. https://doi.org/10.1007/s12559-016-9415-7

De Freitas, L. A., & Vieira, R. (2015, November). Exploring resources for sentiment analysis in Portuguese language. In *2015 Brazilian conference on intelligent systems (BRACIS)* (pp. 152-156). IEEE.

De Hollanda, B. B. B., & Medeiros, J. (2020). Megaeventos esportivos, opinião pública e mídia: um balanço da cobertura midiática e das pesquisas quantitativas sobre os Jogos Olímpicos Rio 2016. *Olimpianos-Journal of Olympic Studies*, *4*, 54-75.

Dickson, T. J., Benson, A. M., & Blackman, D. A. (2011). Developing a framework for evaluating Olympic and Paralympic legacies. *Journal of Sport & Tourism*, *16*(4), 285-302

Dolan, P., Kavetsos, G., Krekel, C., Mavridis, D., Metcalfe, R., Senik, C., ... & Ziebarth, N. R. (2016). The host with the most? The effects of the Olympic Games on happiness. *Centre for Economic Performance*. London School of Economics and Political Science.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018, May). Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 0210-0215). IEEE.

Dayan, D., & Katz, E. (1992). *Media Events: The live broadcasting of history*. Library Quarterly.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: *Pre-training of deep bidirectional transformers for language understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)

Dijk, T. A. V. (2011). *Discourse Studies: A Multidisciplinary Introduction*. SAGE.

Fairclough, N. (2003). *Analysing Discourse: Textual Analysis for Social Research* (1st ed.). Routledge. https://doi.org/10.4324/9780203697078.

Farias, D. S., Matsuno, I. P., Marcacini, R. M., & Rezende, S. O. (2016). Opinion-meter: a framework for aspect-based sentiment analysis. In Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web (pp. 351-354).

Farias, D. H., & Rosso, P. (2017). Irony, sarcasm, and sentiment analysis. In *Sentiment Analysis in Social Networks* (pp. 113-128). Morgan Kaufmann.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82. https://doi.org/10.1145/2436256.2436274.

Felle, T. (2015). Digital watchdogs? Data reporting and the news media's traditional "fourth estate" function. Journalism: Theory, Practice & Criticism, 17(1), 85–96. doi:10.1177/1464884915593246

Fernandes da Silva, C., Guarise Katcipis, L. F., & Zarpellon Mazo, J. (2018). Rio 2016 Olympic Games and immaterial legacy. In Journal of Human Sport and Exercise. https://rua.ua.es/dspace/handle/10045/73308

Fiske, J. (1994) Media Matters. Everyday Culture and Political Change, Minneapolis: Minnesota University Press.

Flyvbjerg, B., Budzier, A., & Lunn, D. (2020). *Regression to the Tail: Why the Olympics Blow Up*. 39.

Fontana, V. (2020). News frames and their influence on political participation and media engagement-a study of the Brazilian Car Wash Operation. *Journal of Latin American Communication Research*, *8*(1-2), 109-129.

Fontes, G. S., Ferracioli, P., & Sampaio, R. (2016). Petrolão na mídia: o enquadramento de 18 meses da Operação Lava Jato nas revistas impressas. *Agenda Política, 4*(3), 238-266.

Foucault, M. (1975). *Surveiller et punir*. Paris, 1, 192-211.

Foucault, M. (1969). *L'archéologie du savoir*. Paris: Gallimard.

Freyre, G. (2019). *Casa-grande & senzala*. Global Editora e Distribuidora Ltda.

Fry, B. (2008). Visualizing data: Exploring and explaining data with the processing environment. " O'Reilly Media, Inc.".

Galdo, R. (2016). Mais de 60% dos cariocas acreditam no sucesso dos Jogos, diz pesquisa. *Globo*. Available on https://oglobo.globo.com/rio/mais-de-60-dos-cariocas-acreditam-no-sucesso-dos-jogos-diz-pe squisa-19635455. Archived on https://web.archive.org/web/20220228152623/https://oglobo.globo.com/rio/mais-de-60-dos-c ariocas-acreditam-no-sucesso-dos-jogos-diz-pesquisa-19635455. Archive date: 15th January, 2022.

Galeano, E. (1994). Fernando Birri, cited by Eduardo Galeano in 'Las palabras andantes?'. *Siglo XXI*.

Garde-Hansen, J. (2011). *Media and memory*. Edinburgh University Press.

Gardiner, B. (2012). Interest in the London Olympics. Available on https://yougov.co.uk/topics/politics/articles-reports/2012/03/30/interest-london-olympics. Archived on https://web.archive.org/web/20220228173751/https://yougov.co.uk/topics/politics/articles-rep orts/2012/03/30/interest-london-olympics. Archive date: 18th January, 2022.

Girginov, V. (2010). *The Olympics: A Critical Reader*. Routledge.

Girginov, V. (2018). *Rethinking Olympic Legacy*. London: Routledge.

Girginov, V., & Preuss, H. (2022). Towards a conceptual definition of intangible Olympic legacy. *International journal of event and festival management*, *13*(1), 1-17.

Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision*. CS224N project report, Stanford.

Gomes, T. G., & Medeiros, C. R. D. O. (2019). Construindo e desconstruindo escândalos de corrupção: a operação Lava-Jato nas interpretações da Veja e Carta Capital. *Organizações & Sociedade*, *26*, 457-485.

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. *Proceedings of the First ACM Conference on Online Social Networks - COSN '13*, 27–38. https://doi.org/10.1145/2512938.2512951.

Graham, S., Weingart, S., and Milligan, I. (2012) "Getting Started with Topic Modeling and MALLET," The Programming Historian 1, https://doi.org/10.46430/phen0017.

Gratton, C., & Preuss, H. (2008). Maximizing Olympic Impacts by Building Up Legacies. *The International Journal of the History of Sport*, *25*(14), 1922–1938. https://doi.org/10.1080/09523360802439023.

Green, K. (2012). London 2012 and sports participation: The myths of legacy. *Significance*, *9*(3), 13-16.

Hall, C. M. (2006). Urban Entrepreneurship, Corporate Interests and Sports Mega-Events: The Thin Policies of Competitiveness within the Hard Outcomes of Neoliberalism. *The Sociological Review*, *54*(2_suppl), 59–70. https://doi.org/10.1111/j.1467-954X.2006.00653.x.

Hammond, C. (2016) Do big sporting events make us do more sport?. BBC. Available on: <https://www.bbc.com/future/article/20160805-do-big-sporting-events-make-us-do-more-sport>. Archived on: <https://web.archive.org/web/20231014193419/https://www.bbc.com/future/article/20160805-do-big-sporting-events-make-us-do-more-sport>. Archive date: 14th October 2023.

Hampton, J. (2012). *The Austerity Olympics: When the Games Came to London in 1948*. Quarto Publishing Group USA.

Han, S. (2020). googletrans 3.0. 0. PyPI Library. Available on https://pypi.org/project/googletrans/. Accessed on 20th January, 2022.

Harvey, D. (2012). *Rebel Cities: From the Right to the City to the Urban Revolution*. Verso.

Haw, S. (2022). The Change in Prevalence of the London 2012 Olympic Legacy Discourse. In Journal of Languages, Texts, and Society, Vol. 5 (2022), 1-23 [online].

Hayes, G., & Karamichas, J. (2012). *Olympic Games, Mega-Events and Civil Societies*. Palgrave Macmillan.

Hedenborg, S. (2015). The Olympic Games in London 2012 from a Swedish media perspective. In *London, Europe and the Olympic Games* (pp. 97-112). Routledge.

Hepp, A., & Couldry, N. (2010). Introduction: *Media events in globalized media cultures*.

Hill, C. R. (1992). *Olympic politics*. Manchester University Press.

Hiller, H. H., & Wanner, R. A. (2018). Public opinion in Olympic cities: From bidding to retrospection. *Urban affairs review*, *54*(5), 962-993.

Honnibal, M., & Montani, I. (2017). spaCy 2: *Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, *30*(4), 330-338.

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software*. PloS one, 9(6), e98679.

Jenkins, R. (2011). *The First London Olympics: 1908: 1908*. Hachette UK.

Kanin, D. B. (2019). *A political history of the Olympic Games*. Routledge.

Karadakis, K., Bopp, T., & Gassman, M. T. (2019). The legacy of the Olympic Games: A viewer's perspective of Olympic Games legacies. *Diagoras: International Academic Journal on Olympic Studies*, *3*, 72-93.

Kassner, N., & Schütze, H. (2019). Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly. *arXiv preprint arXiv*:1911.03343.

Kim, S., Li, F., Lebanon, G., & Essa, I. (n.d.). *Beyond Sentiment: The Manifold of Human Emotions*. 10.

Kitchin, R., Gleeson, J., & Dodge, M. (2013). Unfolding mapping practices: A new epistemology for cartography: *Unfolding mapping practices*. *Transactions of the Institute of British Geographers*, *38*(3), 480–496. https://doi.org/10.1111/j.1475-5661.2012.00540.x.

Kohlschütter, C., Fankhauser, P., & Nejdl, W. (2010). Boilerplate detection using shallow text features. In Proceedings of the third ACM international conference on Web search and data mining (pp. 441-450).

Kovtun (2016). Fake-useragent documentation. Accessed via *https://pypi.org/project/fake-useragent/* on the 28th of July, 2021.

Kumar, E. (2011). *Natural language processing*. IK International Pvt Ltd.

Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News Sharing in Social Media: A Review of Current Research on News Sharing Users, Content, and Networks. *Social Media + Society*, *1*(2). https://doi.org/10.1177/2056305115610141.

Leopkey, B; Parent, M. (2012). Olympic Games Legacy: From General Benefits to Sustainable Long-Term Legacy. The International Journal of the History of Sport, 29(6), 924–943. doi:10.1080/09523367.2011.623006

Lippmann, W. (1922). *Public Opinion*. Macmillan Publishers: New York.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable ai: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Lu, B., Tan, C., Cardie, C., & Tsou, B. K. (2011). Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 320-330).

Ludacer, R. (2018). No one wants to host the Olympics anymore: will they go away? Available on: <businessinsider.com/future-olympics-no-country-wants-to-host-games-2018-2?r=US&IR=T>. Archived on: <https://web.archive.org/web/20230919114838/https://www.businessinsider.com/future-olympics-no-country-wants-to-host-games-2018-2?r=US&IR=T>. Archive date: <19th September 2023>

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30. Available on https://github.com/slundberg/shap. Accessed on 10th January, 2022.

Lutpullaevna, A. F. (2021). Communicative and Pragmatic Features of Colloquial Vocabulary in Newspaper Style. *International Journal on Integrated Education*, *4*(4), 358-362.

MacAloon, J. J. (2008). 'Legacy' as Managerial/Magical Discourse in Contemporary Olympic Affairs. *The International Journal of the History of Sport*, *25*(14), 2060–2071. https://doi.org/10.1080/09523360802439221.

Maingueneau, D. (1984). *Genèses du discours*. Editions Mardaga.

Magalhães, L. F. (2013). As Olimpíadas de 2016: da captação aos possíveis legados socioeconômicos na cidade do Rio de Janeiro.

Maingueneau, D. (2004). *Análise de textos de comunicação*. Cortez.

Maingueneau, D., & Angermuller, J. (2007). Discourse analysis in France. A Conversation. In *Forum Qualitative Research* (Vol. 8, No. 2). Freie Universität Berlin, Institut für Qualitative Forschung.

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. *In Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (p. 28:1–28:5). New York, NY, USA: ACM. https://doi.org/10.1145/3125571.3125585

Maynard, D. G., & Greenwood, M. A. (2014). Who cares about sarcastic tweets? *investigating the impact of sarcasm on sentiment analysis*. In *Lrec 2014 proceedings*. ELRA.

Mannarswamy, S., & Chidambaram, S. (2021). Opening the NLP Blackbox-Analysis and Evaluation of NLP Models: *Methods, Challenges and Opportunities*. In *8th ACM IKDD CODS and 26th COMAD* (pp. 447-448).

McAuley, J., & Leskovec, J. (2013). *Hidden factors and hidden topics: understanding rating dimensions with review text*. In Proceedings of the 7th ACM conference on Recommender systems (pp. 165-172).

McCombs, M. E. (2004). Setting the agenda: *the mass media and public opinion*. Cambridge:Malden, MA: Polity.

Meng, X., Wei, F., Liu, X., Zhou, M., Xu, G., & Wang, H. (2012). Cross-lingual mixture model for sentiment classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 572-581).

Mesquita, F. A., & Bueno, W. da C. (2018). *The Olympic legacy in question: From the conceptual misunderstanding to the negative evaluation of the Brazilian press*. *18*, 14.

Milligan, I. (2019). *History in the age of abundance?: how the web is transforming historical research*. McGill-Queen's University Press.

Millington, R., & Darnell, S. C. (2014). Constructing and contesting the Olympics online: The internet, Rio 2016 and the politics of Brazilian development. *International Review for the Sociology of Sport*, *49*(2), 190–210. https://doi.org/10.1177/1012690212455374.

Misener, L., Darcy, S., Legg, D., & Gilbert, K. (2013). Beyond Olympic legacy: Understanding Paralympic legacy through a thematic analysis. *Journal of sport management*, *27*(4), 329-341.

Misener, L. (2013). A media frames analysis of the legacy discourse for the 2010 Winter Paralympic Games. *Communication & Sport*, *1*(4), 342-364.

Mohammad, S., Salameh, M., & Kiritchenko, S. (2016, May). Sentiment lexicons for Arabic social media. In Proceedings of the tenth international conference on language resources and evaluation (LREC'16) (pp. 33-37).

Molnar, C. *Interpretable machine learning. A Guide for Making Black Box Models Explainable,* 2022. https://christophm.github.io/interpretable-ml-book/.

Moore, S. (2013). A utopian moment of beauty and becoming. In *London 2012: How was it for Us?*. Lawrence & Wishart Ltd.

Moraes, F. (2019). Subjetividade: ferramenta para um jornalismo mais íntegro e integral. *Revista Extraprensa*, *12*(2), 204-219.

Moreno-Ortiz, A. (2017). Lingmotif: Sentiment Analysis for the Digital Humanities. *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 73–76. https://doi.org/10.18653/v1/E17-3019.

Myles, J. F. (2010). *Bourdieu, Language and the Media*. Palgrave Macmillan London. https://doi.org/10.1057/9780230283053.

Nasukawa, T., Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. In: *Proceedings of the 2nd International Conference on Knowledge Capture* (K-CAP 2003), Sanibel Island, FL, USA, 23–25 October 2003. pp. 70–77.

Nauright, J. (2004). Global games: culture, political economy and sport in the globalised world of the 21st century. *Third World Quarterly*, *25*(7), 1325-1336.

Newman, N. (2022). Overview and key findings of the 2022 Digital News Report. Reuters Institute for the Study of Journalism.

Nguyen, D., Vo, K., Pham, D., Nguyen, M., & Quan, T. (2017). A deep architecture for sentiment analysis of news articles. In *International conference on computer science, applied mathematics and applications* (pp. 129-140). Springer, Cham.

Ojala, M. (2021) Is the Age of Impartial Journalism Over? The Neutrality Principle and Audience (Dis)trust. In *Mainstream News, Journalism Studies*, 22:15, 2042-2060, DOI: 10.1080/1461670X.2021.1942150

O'neill, D., & Mulready, M. (2015). The Invisible Woman? A comparative study of women's sports coverage in the UK national press before and after the 2012 Olympic Games. *Journalism practice*, *9*(5), 651-668.

Orlandi, E. P. (2007). *As formas do silêncio: no movimento dos sentidos*. Editora da UNICAMP.

Orlandi, E. P. (2015). Análise de Discurso: princípios e procedimentos. 12ª Edição. *Pontes Editores, Campinas, SP*.

Ou-Yang, L. (2018). Newspaper3k: Article scraping & curation—newspaper 0.0. 2 documentation. Available on https://github.com/codelucas/newspaper. Accessed on 9th September, 2021.

Packard, E. (2021). Declared Insane for Speaking Up: The Dark American History of Silencing Women Through Psychiatry. In TIME. Available on <https://time.com/6074783/psychiatry-history-women-mental-health/>. Archived on <https://web.archive.org/web/20240209225425/https://time.com/6074783/psychiatry-history-women-mental-health/>. Archive date: 9 February 2024.

Padmaja, S., Fatima, S. (2013). Opinion mining and sentiment analysis-an assessment of peoples' belief: a survey. *International Journal of Ad hoc, Sensor & Ubiquitous Computing* 4(1), 21.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford University Press.

Papacharissi, Z. (2016). Affective publics and structures of storytelling: Sentiment, events and mediality. *Information, Communication & Society*, *19*(3), 307–324. https://doi.org/10.1080/1369118X.2015.1109697.

Papanikolaou, P. (2013). Athens 2004. Ten Years Later the Olympic Infrastructure, the Cultural Olympiad and the 'White Elephant' Syndrome. *Journal of Power*, *1*(1), 9.

Pappous, A., & Brown, C. (2018). Paralympic legacies: A critical perspective. *The Palgrave handbook of paralympic studies*, 647-664.

Pêcheux, M. (1969). *Analyse automatique du discours*. Paris: Dunod.

Pereira, D. A. (2021). A survey of sentiment analysis in the Portuguese language. *Artificial Intelligence Review*, *54*(2), 1087-1115.

Piotrkowicz, A., Dimitrova, V., Otterbacher, J., & Markert, K. (2017). The impact of news values and linguistic style on the popularity of headlines on Twitter and Facebook. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 11, No. 1, pp. 767-774).

Pinder, D. (2002). In defence of utopian urbanism: Imagining cities after the 'end of utopia.' *Geografiska Annaler: Series B, Human Geography*, *84*(3–4), 229–241. https://doi.org/10.1111/j.0435-3684.2002.00126.x.

Poerksen, B. (2008). THEORY REVIEW THE IDEAL AND THE MYTH OF OBJECTIVITY: Provocations of constructivist journalism research. *Journalism Studies*, *9*(2), 295–304. https://doi.org/10.1080/14616700701848451.

Poffo, B. (2012). Legados do Pan Rio/2007: uma análise do discurso midiático. *Novas contribuições do LaboMídia/UFSC à pesquisa em Mídia-Educação (Física). Florianópolis: Tribo da Ilha*.

Poynter, G., & MacRury, L. (2012). *Olympic cities: 2012 and the remaking of London*. Ashgate Publishing, Ltd.

Poynter, G (2013). A postcard from Rio. In *London 2012: How was it for us*. London Lawrence & Wishart.

Preuss, H. (2006). *Lasting Effects of Major Sporting Events*. 13. Published on the Internet, www.idrottsforum.org.

Preuss, H. (2015). A framework for identifying the legacies of a mega sport event. *Leisure Studies*, *34*(6), 643–664. https://doi.org/10.1080/02614367.2014.994552.

Preuss, H. (2019). Event legacy framework and measurement. International Journal of Sport Policy and Politics, 103-108. https://doi.org/10.1080/19406940.2018.1490336.

Rayner, P., Wall, P., & Kruger, S. (2001). *Media studies: the essential introduction*. Routledge: New York.

Reis, J., Gonçalves, P., Araújo, M., Pereira, A. C., & Benevenuto, F. (2015, August). Uma abordagem multilíngue para análise de sentimentos. In *Anais do IV Brazilian Workshop on Social Network Analysis and Mining*. SBC.

Richardson, L. (2007). Beautiful soup documentation. April. Accessed via *https://pypi.org/project/beautifulsoup4/* on the 28th of July, 2021.

Rogers, R. (2019). *Doing digital methods*. SAGE Publications Limited.

Rothenbuhler, E.W. (1985) Media Events, Civil Religion, and Social Solidarity: The Living Room Celebration of the Olympic Games, doctoral dissertation, University of Southern California.

Rozado, D. Hughes, R. Halberstadt, J (2022) Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models. PLoS ONE17(10): e0276367. https://doi.org/10.1371/ journal.pone.0276367

Saldaña, Z. W. (2018). Sentiment Analysis for Exploratory Data Analysis. Programming Historian. https://doi.org/10.46430/phen0079.

Scacco, J. M., & Muddiman, A. (2016). Investigating the influence of "clickbait" news headlines. Engaging News Project Report

Schutz, A. (1962). The problem of social reality. In Collected Papers: Volume IV (pp. 71-72). Dordrecht: Springer Netherlands.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2017). Document Level Sentiment Analysis from News Articles. *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1–4. https://doi.org/10.1109/ICCUBEA.2017.8463638

SILVA, M. (2003). Casa-grande & senzala e o mito da democracia racial.: In: SILVA, Mateus Lôbo de Aquino Moura e. Casa-grande & senzala e o mito da democracia racial. ANPOCS, São Paulo. p. 1-24.

Snow, D. A., Vliegenthart, R., & Ketelaars, P. (2018). The Framing Perspective on Social Movements: Its Conceptual Roots and Architecture. In D. A. Snow, S. A. Soule, H. Kriesi, & H. J. McCammon (Eds.), *The Wiley Blackwell Companion to Social Movements* (pp. 392–410). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119168577.ch22.

Souza, F., Nogueira, R., & Lotufo, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In *Brazilian Conference on Intelligent Systems* (pp. 403-417). Springer, Cham.

Steensen, S., & Ahva, L. (2015). Theories of journalism in a digital age: An exploration and introduction. Routledge.

Sugden, A., & Tomlinson, A. (2012). *Watching the Olympics: Politics, power and representation* (1st ed.).

Taj, S., Shaikh, B. B., & Meghji, A. F. (2019). Sentiment analysis of news articles: a lexicon based approach. In *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-5). IEEE.

Tang, T., & Cooper, R. (2013). Olympics Everywhere: Predictors of Multiplatform Media Uses During the 2012 London Olympics. *Mass Communication and Society*, *16*(6), 850–868. https://doi.org/10.1080/15205436.2013.804936.

Tavares, C., Ribeiro, R., & Batista, F. (2021). Sentiment analysis of Portuguese economic news. Sentiment analysis of Portuguese economic news.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the Dimensional and Hierarchical Structure of Affect. *Psychological Science*, *10*(4), 297–303. https://doi.org/10.1111/1467-9280.00157.

Terras, M (2015). Defining Digital Humanities: a reader. Routledge. p.330.

The Olympic Studies Centre (2022). Olympism: From Coubertin to the present day. IOC Library.

Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544–2558.

Thelwall, M., & Kappas, A. (2014). The role of sentiment in the social web. In *Collective emotions* (pp. 375-388). New York, NY: Oxford University Press.

Thelwall, M. (2016). Heart and soul: Sentiment strength detection in the social web with sentistrength. *Cyberemotions: Collective emotions in cyberspace*.

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions* (pp. 119-134). Springer, Cham.

Thomas, R. J., & Antony, M. G. (2015). Competing constructions of British national identity: British newspaper comment on the 2012 Olympics opening ceremony. *Media, Culture & Society*, *37*(3), 493-503.

Tomlinson, A., & Young, C. (2006). *National Identity and Global Sports Events: Culture, Politics and Spectacle in the Olympics and the Football World Cup*. State University of New York Press.

Thomson, A., Schlenker, K., & Schulenkorf, N. (2013). Conceptualising sport event legacy. *Event management*, *17*(2), 111-122.

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science*, *19*(3), 258–273. https://doi.org/10.1177/0963662509102694.

Venturini, T. (2012). Building on faults: How to represent controversies with digital methods. *Public Understanding of Science*, *21*(7), 796–812. https://doi.org/10.1177/0963662510387558.

Venturini, T., Bounegru, L., Jacomy, M., & Gray, J. (2017). 11. How to Tell Stories with Networks Exploring the Narrative Affordances of Graphs with the Iliad. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society* (pp. 155–170). Amsterdam University Press. https://doi.org/10.1515/9789048531011-014.

Venturini, T., Jacomy, M., Bounegru, L., & Gray, J. (2018). Visual Network Exploration for Data Journalists. In S. A. Eldridge & B. Franklin (Eds.), *The Routledge Handbook of Developments in Digital Journalism Studies* (1st ed., pp. 265–283). Routledge. https://doi.org/10.4324/9781315270449-21.

Venturini, T., Jacomy, M., & Jensen, P. (2019). What Do We See When We Look at Networks. An Introduction to Visual Network Analysis and Force-Directed Layouts. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3378438.

Vizeu, A. (2009). Telejournalism as a place of reference and the pedagogical function. Revista FAMECOS, 16(40), 77. https://doi.org/10.15448/1980-3729.2009.40.6321.

Wan, X. (2008). Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 553-561).

Wan, X. (2009). Co-training for cross-lingual sentiment classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (pp. 235-243).

Wei, B., & Pal, C. (2010). Cross lingual adaptation: an experiment on sentiment classifications. In *Proceedings of the ACL 2010 conference short papers* (pp. 258-262).

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse as data: A guide for analysis*. Sage.

White, H. (1980). The value of narrativity in the representation of reality. *Critical inquiry*, *7*(1), 5-27.

Winters, J. (2019). Negotiating the archives of UK web space. In *The Historical Web and Digital Humanities: the case of national web domains*. Routledge.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

Zimbalist, A. (2017). *Rio 2016: Olympic myths, hard realities*. Brookings Institution Press.

Zirin, D. (2014). *Brazil's dance with the devil: The World Cup, the Olympics, and the struggle for democracy*. Haymarket Books.

Appendix

Appendix A: Code to scrape website from list URLs downloaded from SHINE and output TXT file

```
from boilerpipe.extract import Extractor
import os
convertStrings=False
import codecs

## change variables

year = '2006'
domain = '%22guardian.co.uk%22'
domain_name = 'citymonitor'

local = './data/' + domain_name
filename = year + domain + 'search'

with open(local + '/' 'citymonitor.txt') as f:
    urls = f.read().splitlines()

for url in urls:
    if url.startswith('http'):
        extractor = Extractor(extractor='ArticleExtractor', url=url, encoding='latin1')
        extracted_text = extractor.getText()

        cleanQuery = url[27:41]
        filename = cleanQuery
        localfinal = './data/' + domain_name + '/full-articles'
        with open(os.path.join(localfinal, filename + '.txt'), 'w') as file1:
            file1.write(extracted_text)
            file1.close
```

Appendix B: Code to scrape website from Google search and output TXT file

```
import urllib

query = """'site:citymonitor.ai olympic legacy'"""
query = urllib.parse.quote_plus(query) # Format into URL encoding
number_result = 300
domain_name = 'citymonitor2'
local = './data/' + domain_name

import requests
from fake_useragent import UserAgent
```

```
from bs4 import BeautifulSoup

ua = UserAgent()

google_url = "https://www.google.com/search?q=" + query + "&num=" + str(number_result)
response = requests.get(google_url, {"User-Agent": ua.random})
soup = BeautifulSoup(response.text, "html.parser")

result_div = soup.find_all('div', attrs = {'class': 'ZINbbc'})

links = []
titles = []
descriptions = []


for r in result_div:
    # Checks if each element is present, else, raise exception
    try:
        link = r.find('a', href = True)
        title = r.find('div', attrs={'class':'vvjwJb'}).get_text()
        description = r.find('div', attrs={'class':'s3v9rd'}).get_text()


        # Check to make sure everything is present before appending
        if link != '' and title != '' and description != '':
            links.append(link['href'])
            titles.append(title)
            descriptions.append(description)
    # Next loop if one element is not present
    except:
        continue

# Clean titles

new_titles = []

for string in titles:
    new_title = string.replace('?','')
    new_titles.append(new_title)

print(new_titles)

# Clean URLs

import re

to_remove = []
clean_links = []
for i, l in enumerate(links):
    clean = re.search('\/url\?q\=(.*)\&sa',l)

    # Anything that doesn't fit the above pattern will be removed
```

```
    if clean is None:
        to_remove.append(i)
        continue
    clean_links.append(clean.group(1))

# Remove the corresponding titles & descriptions
for x in to_remove:
    del titles[x]
    del descriptions[x]

# Write into TXT file

for url in urls:
    if url.startswith('http'):
        extractor = Extractor(extractor='ArticleExtractor', url=url, encoding='utf8')
        extracted_text = extractor.getText()

        #cleanQuery = url[27:28]
        filename = cleanQuery
        filename = str(startvalue)
        #filename = str(new_title)
        localfinal = './data/' + domain_name + '/full-articles'
        startvalue = startvalue + 1

        with open(os.path.join(localfinal, filename + '.txt'), 'w', encoding='utf8') as file1:
            file1.write(extracted_text)
            file1.close
```

Appendix C: Code to check whether the word legacy is mentioned in the text

```
import os
import re
domain_name = 'g-telegraph'

local = './data/' + domain_name + '/full-articles/'
#localclean = './data/2004/guardian/clean-articles/'
localolympic = './data/' + domain_name + '/olympic-articles/'
items = ['legacy', 'Legacy']

for filename in os.listdir(local):
    if filename.endswith(".txt"):
        with open(os.path.join(local + filename), 'r', encoding='utf-8') as myfile:

            #print(stripTags(myfile.read()))
            content = myfile.read()
```

```
    for i in items:
        lis = re.findall(i, content)
        if len(lis)==0:
            print(filename,'Not found')
        elif len(lis)==1:
            print(filename,'Found once')
            with open(os.path.join(localolympic, filename + '.txt'), 'w', encoding='utf-8') as file1:
                file1.write(content)
        elif len(lis)==2:
            print(filename,'Found twice')
            with open(os.path.join(localolympic, filename + '.txt'), 'w', encoding='utf-8') as file1:
                file1.write(content)
        else:
            print(filename,'Found', len(lis), 'times')
            with open(os.path.join(localolympic, filename + '.txt'), 'w', encoding='utf-8') as file1:
                file1.write(content)
```

Appendix D: Code to retrieve n-grams and word counts

```
import os
import nltk
import wordcloud
import json
import spacy
from spacy.lang.en.stop_words import STOP_WORDS as en_stopwords
from spacy.lang.pt.stop_words import STOP_WORDS as pt_stopwords
from spacy.lang.es.stop_words import STOP_WORDS as es_stopwords
import sys
from nltk.stem import WordNetLemmatizer
from nltk.corpus import wordnet
from nltk import ngrams
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
import pandas as pd
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

def get_stopwords(lang):
    if lang == 'english':
        df_stopwords = set(en_stopwords)
        custom_stopwords = set(['olympics', 'olympic', 'london',
                    'rio', 'legacy', '2012', '2016',
                    'said', 'caption', 'image', 'years',
                    'd','s','t','m','n','ve', 'll', 'xe2', 'x80', 'x99', 'x99s', 'nthe', 'lda', '000', 'xc2'])
        return df_stopwords | custom_stopwords
    elif lang == 'portuguese':
        df_stopwords = set(pt_stopwords)
        custom_stopwords = set(['legado', 'olimpico', 'london',
                    'olímpica', 'londres', 'rio', 'legado',
                    'olímpico', '2012', '2016'])
```

```
      return df_stopwords | custom_stopwords
  def my_tokenizers(doc):
    words = nltk.word_tokenize(doc)
    return [word for word in words if len(word) > 2]

def get_top_n_words(corpus, stopwords, n=20):
    vec = CountVectorizer(stop_words = stopwords).fit(corpus)
    bag_of_words = vec.transform(corpus)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
    return words_freq[:n]

def get_top_n_bigram(corpus, stopwords, n=20):
    vec = CountVectorizer(ngram_range=(2, 2), stop_words = stopwords).fit(corpus)
    bag_of_words = vec.transform(corpus)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
    return words_freq[:n]

def get_top_n_words_tfidf(corpus, stopwords, n=20):
    vec = TfidfVectorizer(stop_words = stopwords).fit(corpus)
    bag_of_words = vec.transform(corpus)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
    return words_freq[:n]

def get_top_n_bigram_tfidf(corpus, stopwords, n=20):
    vec = TfidfVectorizer(ngram_range=(2, 2), stop_words = stopwords).fit(corpus)
    bag_of_words = vec.transform(corpus)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
    return words_freq[:n]

lang = 'english'
directory = 'g-dailymail'

localolympic1 = './data/' + directory1 + '/data_analysis/'

with open(os.path.join(localolympic1 + 'all_texts_together' + directory1 + '.txt'), 'r', encoding= 'utf8') as
myfile1:
    content = myfile1.read()

get_top_n_words([content], get_stopwords(lang), 30)
```

Appendix E: Code to extract entities by combining Spacy with Wikifier

```python
import logging
import spacy
import json
import urllib.parse
from urllib.request import Request
import requests
import time


def get_spacy_annotations(text, language):
        if language == "en":
        spacy_ner = spacy.load("en_core_web_sm")
        elif language == "de":
        spacy_ner = spacy.load("de_core_news_sm")
        elif language == "pt":
        spacy_ner = spacy.load("pt_core_news_sm")
        else:
        logging.error(f"Unsupported language {language}. Please use [en, de, pt]!")
        return []

        doc = spacy_ner(text)
        named_entities = []
        for ent in doc.ents:
        named_entities.append({
        'text': ent.text,
        'type': ent.label_,
        'start': ent.start_char,
        'end': ent.end_char,
        'cms': None,
        })
        return named_entities

def get_wikifier_annotations(text, language):
        threshold = 1.0
        endpoint = 'http://www.wikifier.org/annotate-article'
        language = language
        key = 'oafqnfoihieqdrsxazmvgynivwdddr'
        wikiDataClasses = 'false'
        wikiDataClassIds = 'true'
        includeCosines = 'false'
        data = urllib.parse.urlencode([("text", text), ("lang", language), ("userKey", key),
                        ("pageRankSqThreshold", "%g" % threshold), ("applyPageRankSqThreshold",
"true"),
                        ("nTopDfValuesToIgnore", "200"), ("nWordsToIgnoreFromList", "200"),
                        ("wikiDataClasses", wikiDataClasses), ("wikiDataClassIds", wikiDataClassIds),
                        ("support", "true"), ("ranges", "false"), ("includeCosines", includeCosines),
```

```
                    ("maxMentionEntropy", "3")])

    req = urllib.request.Request(endpoint, data=data.encode("utf8"), method="POST")
    with urllib.request.urlopen(req, timeout=60) as f:
    response = f.read()
    response = json.loads(response.decode("utf8"))
    if 'annotations' in response:
    return {'processed': True, 'annotations': response['annotations']}
    else:
    logging.error(f'No valid response: {response}')
    return {'processed': False, 'annotations': []}

def link_annotations(spacy_annotations, wikifier_annotations):
    POSSIBLE_SPACY_TYPES = ['PER', 'PERSON', 'FAC', 'ORG', 'GPE', 'LOC', 'EVENT', 'MISC']
    linked_entities = []
    for spacy_anno in spacy_annotations:
    # skip all entities with 0 or 1 characters or not in selected spacy types
    if len(spacy_anno['text']) < 2 or spacy_anno['type'] not in POSSIBLE_SPACY_TYPES:
    continue

    related_wikifier_entries = get_related_wikifier_entry(spacy_anno, wikifier_annotations)

    # if no valid wikifier entities were found, try to find entity based on string using <wbsearchentities>
    if len(related_wikifier_entries) < 1:
    # get wikidata id for extrated text string from spaCy NER
    entity_candidates = get_wikidata_entries(entity_string=spacy_anno['text'], limit_entities=1,
language="en")

    # if also no match continue with next entity
    if len(entity_candidates) < 1:
    continue

    # take the first entry in wbsearchentities (most likely one)
    entity_candidate = {
    **{
            'wd_id': entity_candidates[0]['id'],
            'wd_label': entity_candidates[0]['label'],
            'disambiguation': 'wbsearchentities'
    },
    **spacy_anno,
    }
    else:
    highest_PR = -1
    best_wikifier_candidate = related_wikifier_entries[0]
    for related_wikifier_entry in related_wikifier_entries:
    # print(related_wikifier_entry['title'], related_wikifier_entry['pageRank_occurence'])
    if related_wikifier_entry['pageRank_occurence'] > highest_PR:
            best_wikifier_candidate = related_wikifier_entry
            highest_PR = related_wikifier_entry['pageRank_occurence']

    entity_candidate = {
    **{
```

```
                'wd_id': best_wikifier_candidate['wikiDataItemId'],
                'wd_label': best_wikifier_candidate['secTitle'],
                'disambiguation': 'wikifier'
        },
        **spacy_anno,
        }

        linked_entities.append(entity_candidate)

        return linked_entities


def get_related_wikifier_entry(spacy_anno, wikifier_annotations, char_tolerance=2, threshold=1e-4):
        # loop through entities found by wikifier
        aligned_candidates = []
        for wikifier_entity in wikifier_annotations['annotations']:
        if 'secTitle' not in wikifier_entity.keys() or 'wikiDataItemId' not in wikifier_entity.keys():
        continue

        wikifier_entity_occurences = wikifier_entity['support']

        # loop through all occurences of a given entity recognized by wikifier
        for wikifier_entity_occurence in wikifier_entity_occurences:

        if wikifier_entity_occurence['chFrom'] < spacy_anno['start'] - char_tolerance:
        continue

        if wikifier_entity_occurence['chTo'] > spacy_anno['end'] + char_tolerance:
        continue

        # apply very low threshold to get rid of annotation with very low confidence
        if wikifier_entity_occurence['pageRank'] < threshold:
        continue

        aligned_candidates.append({
        **wikifier_entity,
        **{
                'pageRank_occurence': wikifier_entity_occurence['pageRank']
        }
        })

        return aligned_candidates


def get_entity_response(wikidata_id):
        query = """
        prefix schema: <http://schema.org/>
        PREFIX wikibase: <http://wikiba.se/ontology#>
        PREFIX wd: <http://www.wikidata.org/entity/>
        PREFIX wdt: <http://www.wikidata.org/prop/direct/>
        SELECT ?entity ?entityLabel ?entityDescription ?instance ?coordinate ?wikipedia_url ?wdimage
        WHERE {
```

```
        VALUES (?entity) {(wd:%s)}
        OPTIONAL { ?entity wdt:P31 ?instance . }
        OPTIONAL { ?entity wdt:P625 ?coordinate . }
        OPTIONAL { ?entity wdt:P18 ?wdimage . }
        OPTIONAL {
        ?wikipedia_url schema:about ?entity .
        ?wikipedia_url schema:inLanguage "en" .
        ?wikipedia_url schema:isPartOf <https://en.wikipedia.org/> .
        }
        SERVICE wikibase:label {bd:serviceParam wikibase:language "en" .}
        }""" % wikidata_id

        res = get_response("https://query.wikidata.org/sparql", params={'format': 'json', 'query': query})

        if res:
        return res['results']
        else:
        return {'bindings': []}


def fix_entity_types(linked_entities, event_list):
        entity_info = {}

        for i in range(len(linked_entities)):
        wd_id = linked_entities[i]['wd_id']
        if wd_id not in entity_info:
        entity_info[wd_id] = get_entity_response(wikidata_id=wd_id)

        if wd_id in event_list:
        is_event = True
        else:
        is_event = False

        is_person = False
        is_location = False

        information = ["wikipedia_url", "entityDescription", "wdimage"]
        for b in entity_info[wd_id]["bindings"]:
        if "instance" in b and "value" in b["instance"] and b["instance"]["value"].endswith("/Q5"):
        is_person = True

        if "coordinate" in b and "value" in b["coordinate"]:
        is_location = True

        for info_tag in information:
        if info_tag in b and "value" in b[info_tag]:
                linked_entities[i][info_tag] = b[info_tag]["value"]
        else:
                linked_entities[i][info_tag] = ""

        if "wdimage" not in linked_entities[i] or linked_entities[i]["wdimage"] == "":  # set placeholder
image
```

```
        linked_entities[i]["wdimage"] =
"http://www.jennybeaumont.com/wp-content/uploads/2015/03/placeholder.gif"

        # set placeholder for card view
        linked_entities[i]["reference_images"] = [{"url": linked_entities[i]["wdimage"], "source":
"wikidata"}]

        if is_location:
        linked_entities[i]["type"] = "LOCATION"
        if is_person:  # NOTE higher priority if an entity is an instance of person then it cannot be a location
        linked_entities[i]["type"] = "PERSON"
        if is_event:  # NOTE highest priority as the entity is covered by EventKG
        linked_entities[i]["type"] = "EVENT"
        if not (is_location or is_person or is_event):
        linked_entities[i]["type"] = "unknown"

        return linked_entities


def get_wikidata_entries(entity_string, limit_entities=7, language='en'):
        params = {
        'action': 'wbsearchentities',
        'format': 'json',
        'language': language,
        'search': entity_string,
        'limit': limit_entities
        }
        response = get_response('https://www.wikidata.org/w/api.php', params=params)
        if response:
        return response['search']
        else:
        return []


def get_response(url, params):
        i = 0
        try:
        r = requests.get(url, params=params, headers={'User-agent': 'your bot 0.1'})
        return r.json()
        except KeyboardInterrupt:
        raise
        except:
        logging.error(f'Got no response from wikidata. Retry {i}')  # TODO include reason r
        return {}
```

Appendix F: List of most frequently occurring n-grams in the British News Media Corpus, British Government Corpus, and British Activist Corpus; followed by the number of times they occur

Top words all	Top bigrams all	Top words tfidf	Top unigram news	Top bigrams news	Top unigram gov	Top bigrams gov	Top unigram activist	Top bigrams activist
('games', 8386)	('west ham', 1317)	('games', 0.39959768623194253)	('games', 7264)	('west ham', 1446)	('sport', 626)	('uk sport', 217)	('games', 1680)	('stratford city', 254)
('sport', 4977)	('world cup', 556)	('sport', 0.23715689057672049)	('sport', 4955)	('sport england', 475)	('world', 556)	('enter search', 136)	('city', 1287)	('mega event', 212)
('new', 4223)	('long term', 511)	('new', 0.2012283602381938)	('people', 3890)	('world cup', 471)	('uk', 398)	('international inspiration', 119)	('park', 1215)	('lea valley', 202)
('people', 4215)	('young people', 487)	('people', 0.2008471556722678)	('stadium', 3816)	('young people', 457)	('games', 373)	('world class', 108)	('new', 1181)	('clays lane', 180)
('stadium', 3666)	('boris johnson', 433)	('stadium', 0.1746869923355952)	('new', 3314)	('boris johnson', 447)	('international', 319)	('world cup', 97)	('local', 1053)	('new park', 167)
('park', 3356)	('chief executive', 427)	('park', 0.1599153154059622)	('sports', 2577)	('long term', 436)	('events', 273)	('young people', 97)	('housing', 998)	('mega events', 160)
('city', 3122)	('uk sport', 417)	('city', 0.14876508185262635)	('park', 2457)	('chief executive', 421)	('park', 261)	('queen elizabeth', 93)	('development', 734)	('julian cheyne', 153)

('world', 2872)	('paralympic games', 390)	('world', 0.13685243916743847)	('year', 2440)	('lord coe', 402)	('new', 245)	('national lottery', 93)	('people', 732)	('org uk', 136)
('sports', 2566)	('lord coe', 368)	('sports', 0.12227136452076848)	('world', 2297)	('school sport', 305)	('championships', 226)	('elizabeth park', 92)	('area', 697)	('new jobs', 132)
('year', 2496)	('lea valley', 365)	('year', 0.11893582456891587)	('time', 2109)	('prime minister', 300)	('people', 215)	('paralympic games', 80)	('site', 649)	('long term', 127)
('time', 2272)	('sport england', 356)	('time', 0.10826209672298753)	('city', 1960)	('international committee', 297)	('event', 206)	('search press', 68)	('public', 635)	('http www', 124)
('local', 2162)	('international committee', 336)	('local', 0.10302053394150486)	('government', 1947)	('tessa jowell', 265)	('british', 187)	('press enter', 68)	('oda', 629)	('public housing', 122)
('government', 1960)	('stratford city', 319)	('government', 0.09339511865187304)	('west', 1643)	('opening ceremony', 256)	('paralympic', 179)	('world championships', 63)	('urban', 609)	('local people', 121)
('uk', 1927)	('local people', 310)	('uk', 0.09182264981742824)	('bid', 1618)	('premier league', 252)	('sports', 172)	('development corporation', 61)	('event', 579)	('delivery authority', 120)
('public', 1876)	('delivery authority', 290)	('public', 0.08939247070964991)	('like', 1568)	('year old', 247)	('athletes', 170)	('major events', 59)	('barcelona', 523)	('lower lea', 120)

('bid', 1794)	('tessa jowell', 279)	('bid', 0.08548512390890829)	('public', 1507)	('team gb', 238)	('programme', 150)	('lee valley', 54)	('community', 518)	('leabank square', 120)
('east', 1703)	('world class', 270)	('east', 0.0811489219714999)	('east', 1484)	('paralympic games', 230)	('sporting', 146)	('chief executive', 51)	('stratford', 504)	('open space', 115)
('like', 1682)	('prime minister', 267)	('like', 0.08014825998594412)	('ham', 1453)	('uk sport', 216)	('east', 141)	('gold event', 47)	('jobs', 494)	('paralympic games', 113)
('west', 1565)	('year old', 260)	('west', 0.07457314320927617)	('uk', 1373)	('gold medal', 210)	('enter', 138)	('event series', 47)	('regeneration', 492)	('british cycling', 112)
('money', 1562)	('opening ceremony', 230)	('money', 0.07443019149705392)	('money', 1370)	('aquatics centre', 207)	('search', 138)	('share enter', 42)	('land', 472)	('europe 150', 108)
('development', 1555)	('queen elizabeth', 230)	('development', 0.07409663750186866)	('school', 1369)	('new homes', 206)	('host', 136)	('sporting events', 40)	('economic', 451)	('games monitor', 106)
('event', 1515)	('lower lea', 228)	('event', 0.0721906146722386)	('local', 1299)	('co uk', 206)	('development', 134)	('great britain', 37)	('residents', 413)	('urban park', 96)
('housing', 1514)	('team gb', 227)	('housing', 0.07214296410149786)	('team', 1285)	('ken livingstone', 198)	('national', 134)	('children young', 36)	('sector', 411)	('real estate', 94)

('british', 1418)	('premier league', 227)	('british', 0.0675685093 103857)	('centre', 1221)	('local people', 197)	('team', 133)	('east bank', 34)	('facilities', 402)	('jobs created', 94)
('site', 1416)	('mega event', 224)	('site', 0.0674732081 689042)	('britain', 1217)	('delivery authority', 194)	('inspiration', 123)	('high performance', 31)	('report', 400)	('monitor http', 94)
('team', 1402)	('clays lane', 223)	('team', 0.0668061001 7853368)	('football', 1215)	('organising committee', 194)	('young', 122)	('aquatics centre', 29)	('money', 394)	('www gamesmonitor', 94)
('events', 1384)	('new homes', 222)	('events', 0.0659483899 0520015)	('says', 1213)	('sports minister', 193)	('major', 122)	('england hockey', 29)	('mega', 382)	('gamesmonitor org', 94)
('centre', 1366)	('co uk', 221)	('centre', 0.0650906796 3186663)	('british', 1211)	('east end', 190)	('home', 120)	('european championships', 28)	('council', 382)	('arena fields', 92)
('work', 1354)	('school sport', 220)	('work', 0.0645188727 829776)	('going', 1166)	('park company', 189)	('class', 116)	('events uk', 28)	('planning', 381)	('hackney wick', 91)
('way', 1335)	('elizabeth park', 219)	('way', 0.0636135119 3890333)	('way', 1162)	('physical activity', 181)	('year', 116)	('look forward', 27)	('plans', 380)	('health safety', 90)
('going', 1333)	('gold medal', 216)	('going', 0.0635182107 9742181)	('england', 1160)	('development corporation', 179)	('hockey', 113)	('inspiration programme', 27)	('events', 374)	('park europe', 90)
('ham', 1325)	('development agency', 209)	('ham', 0.0631370062 314958)	('work', 1158)	('ham united', 179)	('cup', 110)	('lottery funding', 26)	('sport', 372)	('compulsory

								('purchase', 86)
('athletes', 1319)	('development corporation', 209)	('athletes', 0.06285110280705129)	('athletics', 1150)	('lea valley', 178)	('bid', 110)	('team gb', 24)	('road', 372)	('affordable housing', 85)
('long', 1265)	('ken livingstone', 205)	('long', 0.06027797198705071)	('athletes', 1147)	('playing fields', 176)	('great', 109)	('elite sport', 23)	('plan', 369)	('waltham forest', 84)
('plans', 1260)	('national lottery', 204)	('plans', 0.06003971913334696)	('week', 1137)	('getty images', 176)	('support', 105)	('million children', 23)	('hackney', 367)	('local community', 83)
('community', 1245)	('athletes village', 201)	('community', 0.059324960572235685)	('long', 1113)	('football club', 173)	('success', 101)	('eurohockey championships', 22)	('east', 365)	('athletes village', 82)
('area', 1230)	('hosting games', 198)	('area', 0.05861020201112441)	('funding', 1110)	('david cameron', 172)	('funding', 100)	('british athletes', 22)	('place', 336)	('public funds', 80)
('international', 1203)	('new park', 197)	('international', 0.05732363660112412)	('sporting', 1094)	('mayor boris', 171)	('community', 99)	('women world', 21)	('x9d', 335)	('local residents', 75)
('britain', 1198)	('east end', 196)	('britain', 0.05708538374742036)	('cost', 1056)	('hosting games', 170)	('gold', 97)	('long term', 21)	('use', 325)	('urban renewal', 72)

('funding', 1187)	('million people', 190)	('funding', 0.05656122746927209)	('club', 1040)	('governing bodies', 169)	('queen', 94)	('championships 2015', 21)	('created', 325)	('construction sector', 72)
('sporting', 1186)	('organising committee', 186)	('sporting', 0.05651357689853134)	('need', 1033)	('public money', 168)	('lottery', 94)	('international development', 21)	('time', 319)	('result games', 72)
('facilities', 1181)	('new jobs', 185)	('facilities', 0.05627532404482758)	('good', 1028)	('development agency', 166)	('elizabeth', 93)	('people world', 21)	('uk', 303)	('related mega', 72)
('athletics', 1172)	('mega events', 184)	('athletics', 0.05584646890816082)	('gold', 1009)	('guardian co', 166)	('centre', 89)	('class events', 20)	('way', 298)	('mountain bike', 72)
('cost', 1170)	('aquatics centre', 183)	('cost', 0.05575116776667 9315)	('won', 1006)	('queen elizabeth', 165)	('million', 88)	('commonwealth games', 19)	('like', 297)	('post games', 68)
('school', 1166)	('playing fields', 180)	('school', 0.05556056548371631)	('country', 993)	('elizabeth park', 163)	('hosting', 83)	('local people', 18)	('construction', 295)	('urban governments', 66)
('place', 1149)	('british cycling', 173)	('place', 0.05475050578112353)	('want', 988)	('school sports', 159)	('series', 82)	('loughborough university', 18)	('long', 294)	('media centre', 66)
('need', 1145)	('culture media', 166)	('need', 0.05455990349816053)	('mayor', 986)	('million people', 157)	('place', 81)	('valley hockey', 18)	('year', 286)	('largest new', 66)

('use', 1140)	('park company', 166)	('use', 0.05432165064445677)	('plans', 953)	('world class', 157)	('work', 81)	('copper box', 17)	('centre', 284)	('martin slavin', 65)
('football', 1134)	('getty images', 165)	('football', 0.05403574722001226)	('schools', 951)	('athletes village', 146)	('opportunity', 80)	('box arena', 17)	('communities', 284)	('sports facilities', 63)
('england', 1111)	('guardian co', 165)	('england', 0.05293978409297497)	('great', 947)	('jessica ennis', 144)	('performance', 80)	('inspire generation', 17)	('space', 283)	('international committee', 60)

Appendix G: List of most frequently occurring n-grams in the Brazilian News Media Corpus, Brazilian Government Corpus and Brazilian Activist Corpus; followed by the number of times they occur

Top words all	Top bigrams all	Top words tfidf	Top news	Top bigrams news	Top gov	Top bigrams gov	Top activist	Top bigrams activist
('jogos', 2108)	('jogos olímpicos', 663)	('jogos', 0.3467738993180772)	('jogos', 1456)	('jogos olímpicos', 438)	('esporte', 420)	('ministério esporte', 129)	('jogos', 324)	('vila autódromo', 174)
('esporte', 1413)	('ministério esporte', 406)	('esporte', 0.23244379494138664)	('parque', 1099)	('ministério esporte', 274)	('jogos', 328)	('jogos olímpicos', 92)	('cidade', 316)	('jogos olímpicos', 133)
('parque', 1326)	('parque barra', 303)	('parque', 0.21813196892588727)	('esporte', 974)	('parque barra', 231)	('atletas', 208)	('parque barra', 67)	('vila', 230)	('click here', 39)

('cidade', 1155)	('autoridade governança', 232)	('cidade', 0.19000182813680225)	('cidade', 804)	('autoridade governança', 198)	('brasil', 181)	('alto rendimento', 58)	('moradores', 224)	('here english', 39)
('janeiro', 1031)	('barra tijuca', 227)	('janeiro', 0.169603363471033)	('janeiro', 789)	('barra tijuca', 187)	('ministério', 146)	('confederação brasileira', 37)	('olimpíadas', 214)	('pagina inicial', 38)
('brasil', 929)	('governo federal', 210)	('brasil', 0.15282398124596475)	('prefeitura', 734)	('governo federal', 180)	('centro', 145)	('arena carioca', 35)	('autódromo', 176)	('barra tijuca', 35)
('olímpicos', 862)	('vila autódromo', 205)	('olímpicos', 0.1418022301765572)	('brasil', 695)	('eduardo paes', 161)	('parque', 145)	('autoridade governança', 34)	('olímpicos', 175)	('maria penha', 24)
('prefeitura', 861)	('eduardo paes', 182)	('prefeitura', 0.1416377264292526)	('olimpíada', 630)	('ano passado', 154)	('janeiro', 143)	('últimas notícias', 32)	('favelas', 147)	('theresa williamson', 22)
('atletas', 760)	('centro tênis', 181)	('atletas', 0.12502284795148894)	('ano', 601)	('centro tênis', 149)	('olímpicos', 118)	('centro tênis', 32)	('pessoas', 125)	('eduardo paes', 20)
('ministério', 740)	('ano passado', 169)	('ministério', 0.12173277300539712)	('ministério', 585)	('governança aglo', 145)	('país', 106)	('ministro esporte', 31)	('anos', 121)	('zona norte', 20)
('ano', 735)	('governança aglo', 169)	('ano', 0.12091025426887417)	('olímpicos', 569)	('arena carioca', 125)	('nacional', 99)	('nacional esporte', 31)	('comunidade', 119)	('durante jogos', 20)

('anos', 720)	('arena carioca', 160)	('anos', 0.1184426980593053)	('governo', 557)	('ministério público', 112)	('treinamento', 94)	('controle dopagem', 27)	('prefeitura', 104)
('governo', 661)	('alto rendimento', 159)	('governo', 0.10873697696833445)	('aglo', 553)	('zona oeste', 106)	('arena', 90)	('rede nacional', 26)	('janeiro', 99)
('centro', 659)	('zona oeste', 128)	('centro', 0.10840796947372527)	('anos', 539)	('parque radical', 105)	('foto', 88)	('gov br', 26)	('remoçoes', 90)
('olimpíada', 659)	('confederação brasileira', 122)	('olimpíada', 0.10840796947372527)	('atletas', 535)	('alto rendimento', 100)	('brasileiro', 87)	('leonardo picciani', 26)	('mundo', 87)
('aglo', 638)	('ministério público', 118)	('aglo', 0.10495339078032886)	('milhões', 535)	('iniciativa privada', 88)	('aglo', 85)	('governo federal', 25)	('comunidades', 84)
('milhões', 634)	('parque radical', 115)	('milhões', 0.1042953757911105)	('federal', 515)	('confederação brasileira', 85)	('mundial', 84)	('vôlei praia', 25)	('parque', 82)
('há', 607)	('prefeito eduardo', 97)	('há', 0.09985377461388656)	('há', 512)	('público federal', 84)	('evento', 82)	('nacional treinamento', 24)	('árvores', 82)
('federal', 589)	('centro treinamento', 95)	('federal', 0.09689270716240392)	('centro', 486)	('prefeito eduardo', 83)	('barra', 78)	('esporte alto', 24)	('disse', 71)

('copa mundo', 19)
('plano popular', 19)
('moradores vila', 18)
('zona oeste', 18)
('direitos humanos', 18)
('últimos anos', 17)
('milhões árvores', 17)
('los angeles', 17)

('barra', 581)	('leonardo picciani', 95)	('barra', 0.0955766771839672)	('será', 484)	('arena futuro', 81)	('tênis', 77)	('governança aglo', 24)	('favela', 70)	('comunidades catalisadoras', 16)
('público', 572)	('iniciativa privada', 92)	('público', 0.09409614345822588)	('público', 481)	('rock in', 81)	('crianças', 73)	('canoagem slalom', 22)	('áreas', 69)	('moradores favelas', 16)
('será', 569)	('baía guanabara', 91)	('será', 0.0936026322163 1211)	('obras', 464)	('baía guanabara', 80)	('será', 72)	('centro treinamento', 21)	('durante', 68)	('museu remoções', 15)
('obras', 550)	('ministro esporte', 90)	('obras', 0.0904770610 1752488)	('barra', 461)	('prefeitura janeiro', 79)	('ministro', 70)	('paulo márcio', 21)	('ano', 68)	('sediar jogos', 15)
('disse', 542)	('instalações esportivas', 90)	('disse', 0.0891610310 3908815)	('disse', 433)	('instalações esportivas', 74)	('vôlei', 70)	('durante jogos', 20)	('dia', 65)	('baixa renda', 14)
('dia', 532)	('comitê internacional', 90)	('dia', 0.0875159935 6604225)	('velódromo', 422)	('centro treinamento', 74)	('instalações', 68)	('francisco medeiros', 20)	('zona', 65)	('porto maravilha', 13)
('pessoas', 521)	('prefeitura janeiro', 90)	('pessoas', 0.0857064523 4569176)	('dia', 421)	('instalações olímpicas', 72)	('brasileira', 67)	('universidade federal', 17)	('penha', 65)	('especulação imobiliária', 13)
('arena', 509)	('durante jogos', 89)	('arena', 0.0837324073 7803667)	('arena', 414)	('comitê organizador', 72)	('ano', 66)	('pan americano', 17)	('luta', 64)	('prefeito eduardo', 13)

('evento', 508)	('arena futuro', 88)	('evento', 0.08356790363073208)	('comitê', 402)	('comitê internacional', 71)	('durante', 63)	('tênis mesa', 17)	('há', 63)	('comitê internacional', 13)
('comitê', 480)	('canoagem slalom', 87)	('comitê', 0.07896179870620354)	('presidente', 399)	('leonardo picciani', 68)	('desenvolvimento', 62)	('foto francisco', 17)	('cidades', 62)	('políticas públicas', 12)
('velódromo', 470)	('rock in', 87)	('velódromo', 0.07731676123315763)	('evento', 381)	('corpo bombeiros', 67)	('alto', 61)	('jiu jitsu', 17)	('locais', 62)	('planejamento urbano', 12)
('projeto', 469)	('público federal', 87)	('projeto', 0.07715225748585304)	('arenas', 370)	('bbc brasil', 66)	('modalidades', 60)	('notícias tweet', 16)	('governo', 60)	('https co', 12)
('país', 464)	('instalações olímpicas', 81)	('país', 0.07632973874933008)	('pessoas', 361)	('zona sul', 65)	('anos', 60)	('tiro esportivo', 16)	('remoção', 59)	('feira jul', 12)
('presidente', 463)	('comitê organizador', 79)	('presidente', 0.07616523500202549)	('instalações', 359)	('radical deodoro', 65)	('rendimento', 59)	('ricardo leyser', 15)	('público', 57)	('zona sul', 12)
('olimpíadas', 454)	('zona sul', 77)	('olimpíadas', 0.07468470127628418)	('projeto', 358)	('canoagem slalom', 62)	('estrutura', 59)	('paralímpico brasileiro', 15)	('região', 57)	('mídia internacional', 12)
('instalações', 442)	('paulo márcio', 77)	('instalações', 0.07271065630862909)	('paes', 345)	('ministro esporte', 58)	('federal', 58)	('esporte brasileiro', 15)	('tóquio', 57)	('linhas ônibus', 12)

('durante', 421)	('copa mundo', 71)	('durante', 0.06925607761523268)	('país', 333)	('arenas cariocas', 58)	('eventos', 58)	('esporte leonardo', 15)	('segurança', 56)	('cidades sede', 12)
('foto', 411)	('vôlei praia', 71)	('foto', 0.06761104014218677)	('prefeito', 323)	('velódromo parque', 57)	('modalidade', 56)	('circuito mundial', 15)	('projeto', 55)	('dona penha', 12)
('mundo', 411)	('olímpicos paralímpicos', 71)	('mundo', 0.06761104014218677)	('gestão', 320)	('marcelo crivella', 56)	('competições', 56)	('atletas brasileiros', 14)	('casa', 55)	('andrew zimbalist', 12)
('vila', 400)	('zona norte', 69)	('vila', 0.06580149892183627)	('foto', 315)	('olímpicos paralímpicos', 56)	('projeto', 56)	('atletas alto', 14)	('internacional', 54)	('ano olimpíadas', 11)
('arenas', 399)	('radical deodoro', 69)	('arenas', 0.06563699517453168)	('feira', 303)	('paulo márcio', 56)	('milhões', 55)	('olímpicos paralímpicos', 14)	('história', 54)	('baía guanabara', 11)
('paes', 383)	('corpo bombeiros', 67)	('paes', 0.06300493521765824)	('hoje', 302)	('cidade janeiro', 55)	('carioca', 53)	('rededoesporte gov', 14)	('casas', 54)	('inicial olhonasremoções', 10)
('hoje', 379)	('comitê brasil', 66)	('hoje', 0.06234692022843987)	('durante', 290)	('comitê brasil', 53)	('pista', 52)	('secretaria especial', 13)	('construção', 53)	('complexo alemão', 10)
('eventos', 375)	('cidade janeiro', 66)	('eventos', 0.06168890523922151)	('eventos', 287)	('velódromo centro', 52)	('judô', 52)	('george hilton', 13)	('brasil', 53)	('região portuária', 10)
('prefeito', 374)	('bbc brasil', 66)	('prefeito', 0.06152440149191692)	('tênis', 286)	('esporte sediou', 52)	('presidente', 49)	('bolsa atleta', 13)	('famílias', 51)	('cidade janeiro', 10)

('tênis', 363)	('velódromo parque', 65)	('tênis', 0.059714860271566424)	('plano', 283)	('ministério cidadania', 51)	('social', 49)	('lei incentivo', 13)	('resistência', 51)	('longo prazo', 10)
('gestão', 360)	('arenas cariocas', 64)	('gestão', 0.05922134902965265)	('mundo', 279)	('campo golfe', 51)	('paulo', 49)	('comitê paralímpico', 13)	('projetos', 51)	('ano jogos', 10)
('sendo', 357)	('marcelo crivella', 61)	('sendo', 0.05872783778773888)	('paulo', 276)	('tribunal contas', 50)	('internacional', 49)	('desenvolvimento esporte', 13)	('plano', 50)	('24 milhões', 10)
('feira', 353)	('nacional esporte', 60)	('feira', 0.058069822798520514)	('sendo', 272)	('autoridade pública', 50)	('tweet', 49)	('preparação atletas', 13)	('ônibus', 50)	('nolympicsla', 10)
('plano', 353)	('últimos anos', 59)	('plano', 0.058069822798520514)	('autoridade', 265)	('zona norte', 49)	('competição', 48)	('comitê brasil', 13)	('sido', 49)	('brt transcarioca', 10)
('carioca', 350)	('campo golfe', 58)	('carioca', 0.057576311556606745)	('carioca', 265)	('linha metrô', 49)	('atleta', 48)	('centros treinamento', 13)	('sociais', 48)	('heloisa helena', 10)