Get Brexit Done: A Comparative Analysis of the Political Discourse during this Process

Construyendo el Brexit: un análisis comparativo del discurso político durante dicho proceso

Aroa ORREQUIA-BAREA

aroa.orrequia@uca.es

University of Cádiz

ABSTRACT

The Brexit process started on 23rd June 2016 when a referendum was held to vote whether the UK was leaving the EU or not. However, it did not become a reality until 31st January 2020, when the UK officially left the EU. Many debates have taken place to reach this agreement between the most influential politicians in the country. The main objective of this paper is to analyse the political discourse of the two main protagonists of this process: Boris Johnson, the Prime Minister, and Jeremy Corbyn, the leader of the opposition. The analysis is twofold: on the one hand, a linguistic analysis was carried out to compare the word choice of each politician; on the other, Sentiment Analysis techniques were applied to explore the general polarity of the political discourse.

KEYWORDS

Corpus Linguistics, Critical Discourse Analysis, Political Discourse, Sentiment Analysis.

RESUMEN

El Brexit comenzó el 23 de junio de 2016 cuando tuvo lugar un referéndum para votar si el Reino Unido abandonaría la Unión Europea o no. Sin embargo, no fue hasta el 31 de enero de 2020 cuando se hizo realidad y el Reino Unido oficialmente dejó de pertenecer a la Unión Europea. Numerosos debates han tenido lugar para alcanzar este acuerdo entre los políticos más importante del país. El principal objetivo de este artículo es analizar el discurso político de los dos protagonistas de este proceso: Boris Johnson, Primer Ministro, y Jeremy Corbyn, líder de la oposición. El análisis tiene un doble objetivo: por un lado, se trata de un análisis léxico para comparar la elección de la palabra de cada uno de los políticos; por otro, se ha realizado análisis de sentimiento para ver la polaridad general del discurso político.

PALABRAS CLAVE

Brexit, Lingüística de corpus, análisis crítico del discurso, discurso político, análisis de sentimiento .

1. INTRODUCTION

Since 23rd June 2016, when a referendum was held in the United Kingdom (UK) to vote whether the country would leave or not the European Union (EU), many political debates and speeches have taken place so as to defend both stances: to remain or to leave the EU. *Brexit* is the name that this process receives as it is a blend word from *Britain* and *exit*.

The EU is a political and economic union of 28 states of Europe, which allows the free movement of people and trade among these countries. Although the UK has belonged to the EU since 1973, this country has not always accepted European policies. As its primary motivation was economic, Britain has not taken part in numerous political or integration policies. For example, it was negotiated to be excluded from the Maastricht Treaty in 1992, not adopting the euro's common currency. Similarly, the British parliament did not accept belonging to the Schengen Area when it was proposed back in 1985. These actions revealed a not fully committed country to the EU. In fact, the UK's stance on the EU has been distant and the British parliament has always had a Eurosceptic discourse.

However, it was not until the 2014 elections and the rise of the UKIP, the UK Independence party, when the independence issue gained importance in society. Although this idea was already in the British's minds, the Brexit process did not start until 23rd June 2016 when David Cameron, Prime Minister, called a referendum to ask the population whether they wanted to leave the European Union. Unexpectedly, 52% of the population voted in favour of leaving the EU, whereas only 48% voted against it. Therefore, and as the people wanted, negotiations started in order to leave the community. It was a complicated process since there was no precedent of a country leaving the EU. In the same vein, the referendum's success came as a surprise for the Conservative party, as they had not drawn up a clear plan to leave.

After the referendum, David Cameron had to resign due to the lack of support in such negotiations. On 13th July of the same year, Theresa May became his successor, but she also left three years later because she could not reach an agreement for the Brexit process. On 24th July 2019, Boris Johnson, from the Conservative Party, was appointed Prime Minister of the British Parliament. It was him who finally got the aim of leaving the EU, as he promised during the election campaign. The main reason he gave to his voters was that they needed to take control of their own affairs, mainly in issues such as emigration and borders. In fact, *Get Brexit done* was his slogan in every political speech he delivered during this period. Finally, on 31st January 2020, Brexit became a reality and the UK officially left the EU. In fact, this day has been called 'Brexit Day'. However, it was not until January 2021 when the UK definitively left the European Union.

During all this time, Jeremy Corbyn, leader of the Labour Party and the opposition, has stood up for the remaining of the UK in the EU. Through his speeches and debates, he has made clear that he was against Brexit and tried to persuade people not to leave the EU. However, the election results led him to resign.

This paper aims to analyse the political discourse during the Brexit process, mainly to compare the two different stances on the issue: in favour, Johnson, and against, Corbyn. The analysis is twofold: firstly, a linguistic analysis will be performed to compare the lexicon and the word choice used by both politicians in their speeches. Secondly, Sentiment Analysis techniques will be applied to study the polarity of the discourse of each candidate.

The paper is organised as follows. Section 2 offers a review of the theoretical background, describing the branches of Critical Discourse Analysis and Sentiment Analysis. Section 3 describes the methodology followed in compiling the corpus, which is detailed in Section 4. In Section 5, we discuss the most relevant results. Finally, Section 6 provides the most important conclusions and Section 7 provides the bibliographical references used in this research.

2. THEORETICAL BACKGROUND

2.1. Critical discourse analysis

Communication, in general, and language, in particular, are essential elements to do politics. In fact, language has a vital role in political discourse. Language is usually influenced by the speaker's ideology, role, cultural background or social status, but when using language, such characteristics can also be depicted (Ekawati, 2019, p. 5). However, when analysing political speeches, it is not sufficient to only pay attention to the text. Other external elements and relations also need to be taken into account, such as power. According to Fairclough (2013), power and discourse somehow *flow into* each other since power can be practised using speech, but, at the same time, power relations are reflected in discourse (p. 4). Thus, language itself can reflect or create power and this is achieved through language use (Ekawati, 2019, p. 5).

Critical discourse analysis (CDA) studies the discourse and the social relations of all the elements that take part in the process of communication, such as the situation, the participants, the institutions, the social structure and any other object that may be present. Hence, discourse is seen as a form of social practice and as such, CDA cannot just analyse the discourse, but the relations between discourse and other elements of the social process should be taken into account. In fact, to understand the linguistic expressions used in the discourse, it is necessary to understand the context in which it occurs (Ekawati, 2019, p. 7). In words of Fairclough and Wodak (1997), "discourse is socially constitutive as well as socially conditioned" (p. 258).

 CDA is considered an interdisciplinary way of analysis since different disciplines are taken into account, such as linguistics, politics or sociology, among others (Fairclough, 2013, p. 4). According to Weiss and Wodak (2007), "language is not powerful on its own" (p. 14), but it acquires power when influential people use it with such a purpose. For this reason, it is necessary an interdisciplinary work to understand how language works when conveying knowledge or exercising power (Weiss & Wodak, 2007, p. 14). The main aim of this type of analysis is "to produce interpretations and explanations of areas of social life which both identify the causes of social wrongs and pro-

duce knowledge which could […] contribute to righting or mitigating them" (Fairclough, 2013, p. 8).

As CDA is an interdisciplinary work, there is not just one methodology to carry out the analysis, but different methods and theoretical backgrounds oriented towards different data can be used. One of the most frequently used methodologies to perform CDA is Corpus Linguistics. This discipline studies collections of texts that occur in natural language and that are usually stored in computers (Baker, 2006, p. 1). Corpus Linguistics analyses the language based on natural language samples in use (McEnery & Wilson, 1996, p. 1 as cited in Baker, 2006). According to Lee (2008), being a corpus linguist is not just about studying the language but also about having a set of techniques that helps the linguist to investigate such amount of information (p. 88).

In the field of CDA, corpus analysis can reveal traces of ideologies in the discourse as well as uncover the intentions that the participants may have when using the language (Baker, 2006, p. 13). The results obtained in the corpus analysis need to be interpreted, taking into account the social factors and the context which surrounds that particular text. At this point, CDA plays an essential role since explanations of the results and patterns discovered in the analysis need to be provided. Additionally, when analysing a text, it is not only important what is said in the text but also what is not said since the absence of particular discourse features also implies an intention (Baker, 2006, pp. 18-19).

2.2. Sentiment analysis

Sentiment Analysis (SA) is one of the applications of Natural Language Processing (NLP) systems. NLP deals with the development of programs or applications that emulate human language behaviour, that is, all the activities that humans do when understanding or producing language. The central task of NLP is to create applications that can be useful for society and SA is one of them.

SA, also called Opinion Mining, studies people's opinions or attitudes towards particular objects, events or topics and their aspects (Liu, 2012). It is a relatively new field since it was not until 2000 with the development of the web 2.0 when researchers started to investigate it. Unlike the Web 1.0, where users only could read and watch the web content, in the Web 2.0, users can also create content, that is, to generate information that can be published on the net. This new web type is also called 'collaborative website', since users can interact among them writing blogs, leaving opinions or reviews, posting comments or tweeting. It is not surprising then that the rise of SA has coincided with that of social networks (Liu, 2012, p. 7).

SA techniques are generally applied to fields where subjectivity is an essential aspect, such as users' opinions on the Internet. Traditionally, opinionated content has been related to users' reviews and posting on social media, but it is actually more than that. Nowadays, there are different domains where sentiment analysis techniques can be applied, such as the financial, healthcare, political or social contexts. In fact, in a survey conducted in the US, over 2,500 Americans, 31% of the participants declared that the Internet was their source of information in the 2006 political campaign (Pang & Lee, 2008, p. 2). Additionally, posting on social networks can have an impact on so-

ciety as it happened with demonstrations such as the ones in the Arab countries in 2011 (Liu, 2012, p. 9).

SA studies the words, phrases and idioms of a text which carry positive or negative sentiments. There exist two main approaches to do this: the *machine learning* approach; and the lexical method. The former trains algorithms using a tagged dataset and compare their performance. This process is also known as the *black box method*. The latter uses annotated lexicons in which the polarity is assigned to each word. The most common lexicons identify words in terms of their polarity, classifying them in positive and negative, and sometimes even neutral. Examples of these lexicons are the NRC lexicon (Mohammad & Turney, 2013) or the BING lexicon (Hu & Liu, 2004), which will be used in this paper to analyse the texts.

SA can also be used in political discourse or debates since the candidates give their perspective about current affairs (Pang & Lee, 2008). When applying SA to politics, two different types of research can be done. On the one hand, research can be focused on what voters think, as can be seen in Laver et al. (2003), where political discourse is analysed to extract political stances. On the other hand, research can be helpful to know politicians' stances and thus, improve the quality of the information to what voters can access. For example, in a study, Thomas, Pang & Lee (2006) analyse the debates of the US Congress to determine which discourses were in favour of and which ones were against a specific policy (Pang & Lee, 2008). In this sense, the political field in general and the political debates, in particular, are interesting scenarios to analyse, as candidates are normally in opposition and their opinions can be compared or contrasted (Fernández Melendres & Moreno-Ortiz, 2020).

3. METHODOLOGY

In this section, we discuss the process of compilation of the political speeches. We have followed two main steps: the collection of the texts and processing them. These steps will be explained in what follows.

3.1. Compilation of the texts

The texts that will be analysed in this work consists of the most relevant political speeches on behalf of the two most influential politicians in the UK during the debate on Brexit: Boris Johnson, current PM from the Conservative Party; and Jeremy Corbyn, leader of the opposition and the Labour Party.

The corpus consists of a compilation of Johnson's and Corbyn's political speeches delivered over the period from 24th July 2019 to 31st January 2020, coinciding with the moment when Johnson was appointed PM until the day when the emancipation process was finished. Although Corbyn has been defending his remain stance from the beginning of this process, we decided to choose only the ones delivered in this period to compare them with Johnson's, who was the one to carry it out.

The political speeches have been extracted from the website, particularly from two differ-

ent web pages. On the one hand, we have retrieved Johnson's speeches from the British Govern-ment webpage[1], which compiles all the speeches delivered by the PM. This webpage offers a vast amount of information about everything related to the government: pieces of news, press releases, new policies, among others. This is the reason why we needed to restrict our search using some fil-ters. In the *government* section, where the speeches delivered by members of the government can be found, we have used the search box to look for some keywords that generally appear in the headlines of the speeches: *PM* –Prime Minister– and *speech*[2]. Afterwards, we have applied another filter to order the speeches chronologically and we have set the period from 24th July 2019 until 31st January 2020. From this selection, we have only chosen the speeches corresponding to the PM. Finally, we have checked whether the speeches talked about *Brexit* or not. We have compiled a total of 10 speeches which comprises a corpus of 16,834 words. The dates can be seen in Table 1.

Date	Speech title[3]
24th July 2019	Boris Johnson's first speech as Prime Minister
25th July 2019	PM statements on priorities for the government
27th July 2019	PM speech at Manchester Science and Industry Museum
2nd September 2019	Prime Minister's statement
3rd September 2019	PM statement on G7 Summit
17th October 2019	PM press conference at EU Council
19th October 2019	PM statement in the House of Commons
13th December 2019	PM statement in Downing Street
31st December 2019	Prime Minister's New Year's message
31st January 2020	PM address to the nation

Table 1. Boris Johnson's political speeches.

On the other hand, the official webpage of the Labour Party compiles the speeches and press releases delivered by the leader, so we extracted Corbyn's political speeches from here. In this webpage, there is a search box where we typed the politician's name to retrieve all the speeches chronologically. Additionally, we have applied the same search criterion, that is, to check whether the word *Brexit* appears or not on the text. As a result, we have compiled a total of 28 speeches which comprises a corpus of 34,657 words. The dates can be seen in Table 2.

---

[1] Available at: https://www.gov.uk/.
[2] We also tried with other relevant words such as the PM's name, *Boris Johnson*, and *statement* but the results were the same.
[3] As can be seen on the webpage.
[4] Available at: https://labour.org.uk/.

Date	Speech title
25th July 2019	Jeremy Corbyn responds to Boris Johnson's first statement in the House as Prime Minister
19th August 2019	Jeremy Corbyn speech in Corby today
2nd September 2019	Jeremy Corbyn speech in Salford today
3rd September 2019	Jeremy Corbyn responding to the Prime Minister Boris Johnson in the House today
3rd September 2019	Jeremy Corbyn's emergency debate statement
10th September 2019	Jeremy Corbyn speech to TUC Congress
25th September 2019	Jeremy Corbyn's response to Boris Johnson in Parliament
3rd October 2019	Jeremy Corbyn responds to the Prime Minister's Brexit statement
10th October 2019	Full text of Jeremy Corbyn's speech in Northampton
17th October 2019	Jeremy Corbyn responds to Brexit Deal
19th October 2019	Jeremy Corbyn's response to the Prime Minister's statement
22nd October 2019	Jeremy Corbyn's speech on the Withdrawal Agreement Bill 2nd reading
29th October 2019	Jeremy Corbyn's early election statement
31st October 2019	Jeremy Corbyn's first major speech of the General Election Campaign
5th November 2019	Jeremy Corbyn's Brexit speech in Harlow
6th November 2019	Jeremy Corbyn's speech in Telford on leadership and what a Labour government will achieve
18th November 2019	Jeremy Corbyn's speech at the Confederation of British Industry
24th November 2019	"Boris Johnson has launched a manifesto for billionaires" Jeremy Corbyn responds to the Conservative manifesto
26th November 2019	"Please take five minutes and register to vote now" – Jeremy Corbyn on the final day of voter registration
27th November 2019	Jeremy Corbyn reveals unredacted documents about secret US-UK trade talks
28th November 2019	Full text of Jeremy Corbyn's speech on Labour's environmental policies
1st December 2019	Full text of Jeremy Corbyn's speech in York
6th December 2019	Jeremy Corbyn reveals leaked Brexit documents
7th December 2019	On 12th December, people have the chance to vote for the most ambitious plan to transform our country in decades – Jeremy Corbyn
10th December 2019	Corbyn's message to undecided voters: vote for hope and a Labour government that will be on your side
11th December 2019	"Vote Labour and get a government that can bring hope to the future" – Jeremy Corbyn speaking on the campaign trail
20th December 2019	Jeremy Corbyn's response to the EU Withdrawal Agreement Bill
30th January 2020	Jeremy Corbyn comment on the UK's exit from the EU

Table 2. Jeremy Corbyn's political speeches.

The scrapping technique has been used to compile such texts. It consists of the automatic downloading of texts. A scraper, or a spider, as it is also known, is an application that visits

websites in order to extract the content that the user wants (Gerdes & Stringam, 2008, p. 310). There are different tools to carry out this task, such as Scrapy, but we will use the same programming language as to analyse this work. Therefore, the texts will be retrieved in plain text format using the programming language R to make the analysis easier. It is necessary to inspect the source code or HTML code to download the content. As we will download specific information, namely the texts of the speeches, we need to specify where this information is contained so that R can extract it. The corpus contains 51,491 words and its distribution can be seen in Table 3.

Politician	Tokens	Types
Boris Johnson	16,834	2,846
Jeremy Corbyn	34,657	3,973

Table 3. Distribution of the two corpora.

As can be seen, we found more speeches containing the word *Brexit* on the labourist webpage than in the PM's one. Therefore, Corbyn's corpus is bigger than Johnson's. However, the fact that they are not balanced will not affect the results since we will use the normalised frequency –per 10,000 words– to compare the corpora.

3.2. Processing

The text processing is carried out using different tools to allow the user to and analyse the texts. This process will have two main stages: firstly, the text will be analysed using Corpus Linguistics statistics, and secondly, the text will be analysed in terms of SA. Thus, we will need four main tools to carry out the processing and the analysis: the programming language R, the libraries, the lexicons and Sketch Engine.

R is a high-level programming language commonly used in Statistics and Mathematics that can also be applied to other sciences, such as Language (Desagulier, 2017). RStudio is frequently used together with R. It is a free software that works as a desktop to manage R easily. Although there exist other programming languages, we will use R because of three main advantages. Firstly, it analyses texts in an automatic and computational way. According to Fradejas Rueda (2019), "the automatic analysis of texts […] is a set of research techniques that use computational analysis to reveal patterns in texts". Secondly, apart from the statistics, R offers a wide range of possibilities in terms of data visualisation. Last but not least, R is a free open access language programming, so there is a huge community developing new libraries and applications that increase the functions of R (Fradejas Rueda, 2019). In fact, there are 16,215 libraries, also called packages, which are available on the web.

According to Santana & Nieves Hernández, "a package is a collection of functions, data and R code that is stored in a folder with a well-defined structure and which is easily accessible by R" (2016, as cited in Murillo Lanza, 2017, p. 64). The use of these libraries allows the user to perform more complex analysis and increase R's functions. In the linguistic field, for example, many li-

braries have been developed to analyse texts. In this research, we have used the packages shown in Table 4.

Packages	Uses
rvest, RCurl, XML	To extract the data from the Internet.
tidydata environment: -tidytext -tidyverse	To do a corpus-based analysis: -They transform the texts into tables so that R can interpret and analyse them. -They divide the text into words or paragraphs, count words or remove stop-words. -Tidyverse includes the SA function.
Udpipe	To tag the text.
ggplot, scales	To visualise the results.
syuzhet (Jockers, 2015)	To do Sentiment Analysis.

Table 4. Summary of the packages used and their functions in this research.

As previously said, the task of SA systems is to identify the words in the text and determine their polarity, taking into account the information provided in the lexicon (Moreno-Ortiz, 7-11 de septiembre de 2017, p. 133). An example of this type of lexicon is the BING lexicon (Hu & Liu, 2004), which has been used to do this work. This dictionary classifies words in positive and negative. There are also some lexicons based on emotions, as is the case of Word Affect Intensities by Mohammad (2019), which assigns four basic emotions to words: anger, fear, sadness and joy. However, in this research, we will use a mixed lexicon since it combines the polarity of words –positive and negative– with the emotions. This lexicon, developed by Mohammad and Turney (2013), is called NRC Word-Emotion Association Lexicon (EmoLex) and comprises 14,182 word which contains two sentiments: positive and negative; and eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. Although this lexicon was initially elaborated in English, there are versions in other languages. In this case, as the political speeches are in English, we will use the original version.

Although the programming language R provides a comprehensive analysis of the texts, we decided to use Sketch Engine not only to enrich the analysis but mainly to see the words in contexts. Sketch Engine[5] is a Corpus Linguistics software that has two interesting functions: it allows the user to create their own corpora, either if they have been extracted from the web or include their own corpora; and the software contains a significant number of corpora, such as the British National Corpus which allows the user to compare with their own. One of the most powerful functions of Sketch Engine is the so-called Concordance, as we can see the word in the context in which it occurs (Kilgarriff *et al.*, 2014). This is why we have used it in our analysis, as sometimes context is funda-

---

[5] Available at: https://www.sketchengine.eu/.

mental to reach some conclusions about a word.

4. ANALYSIS OF THE RESULTS

A twofold analysis will be presented in this section. Firstly, we will analyse and compare the word choice of the two politicians. Secondly, we will focus on the SA of the speeches. A mixed-method corpus qualitative approach has been used in this work as we provide not only results but also their interpretation within the particular context.

4.1. Lexical comparison

Language is a fundamental element in the political field. Therefore, politicians plan their speeches very carefully. They are even helped by experts because political discourse is the only product they can give their voters and the tool they have to convince the population. Therefore, political candidates cannot afford to make any mistake when choosing a word that can be offensive or that may not include the whole population. If so, they will drag themselves into the spotlight. For this reason, we consider that a lexical comparison of the word choice of each politician is essential in the discourse analysis of the speeches. This analysis will reveal, as expected, ideological traces in each politician's discourse. Figure 1 shows the relative frequency of the most frequent words corresponding to each candidate's corpus.
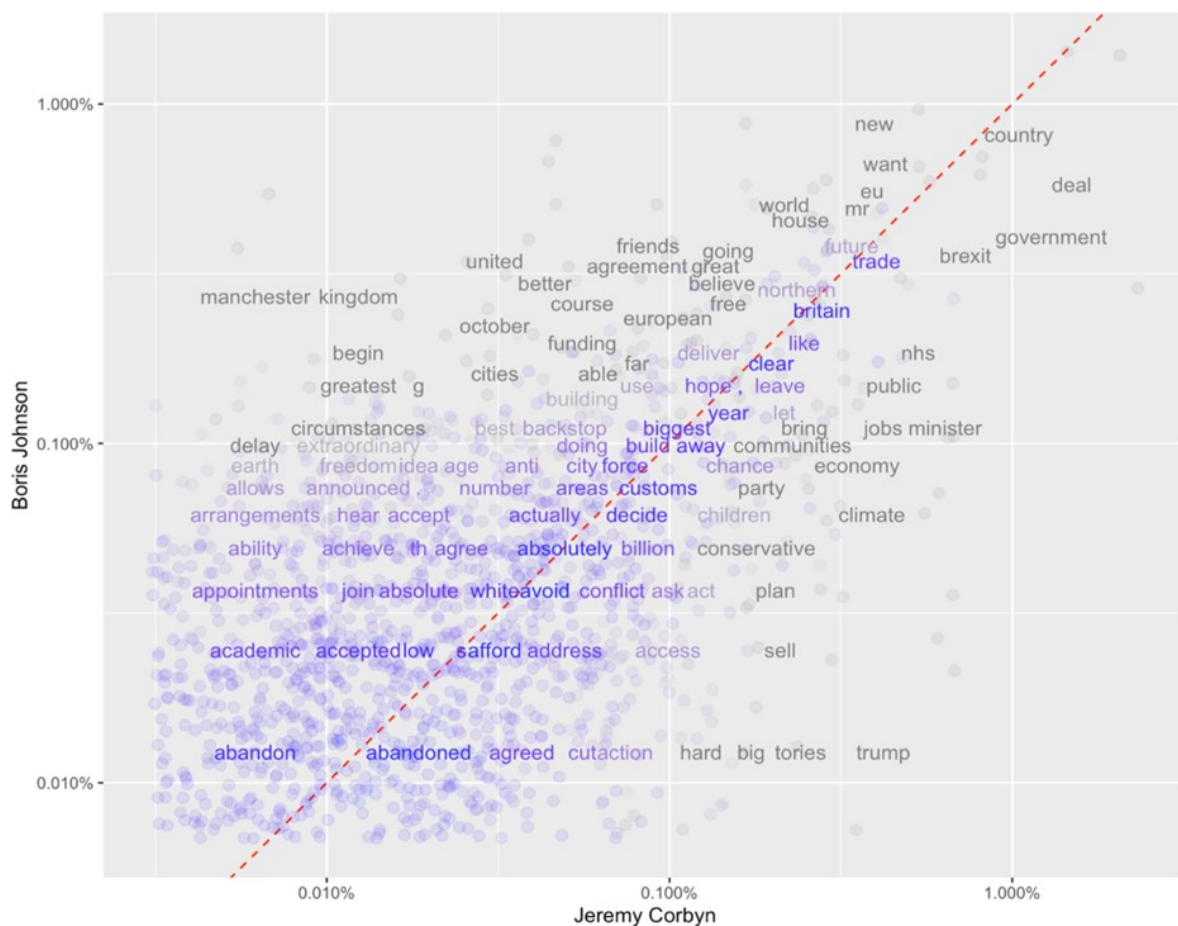

Figure 1. Relative frequency of the most frequent words in Johnson's and Corbyn's speeches
.

As can be seen, there are some items which are shared by both politicians, such as *Britain*, *country, government* or *trade*. It is not surprising to find words like these as they are related to the political activity and their own country. However, there are some of them which are more frequent in one corpus than in the other. For example, *eu*, to refer to the European Union, is present in both corpora. Nevertheless, as expected, Johnson used it more frequently —21.3— than Corbyn —12.6—. The same thing happens with *Brexit*, as it unexpectedly appears twice more in Corbyn's corpus —34— than in Johnson's —17.2—.

Deal is another interesting word to analyse as it is also present in both corpora. In this case, this item is more frequent in Corbyn's corpus —71.8— than in Johnson's —35—. However, *deal* is used with two different purposes in each corpus. Regarding Corbyn's speeches, in most cases, this word appears together with *no* to criticise the *No Deal* policy. The labourist candidate refers to the fact that the UK may leave the EU without reaching an agreement, as shown in (1) and (2):

> "But first, we face the threat of No Deal which would decimate industry and destroy people's jobs in those very same regions" —2nd September 2019. Jeremy Corbyn speech in Salford today—.
> "There can be no doubt about the damage No Deal will do to our economy" —2nd September 2019. Jeremy Corbyn speech in Salford today—.

On the other hand, Johnson indicates the opposite as he promised their voters that they would get a deal to leave the EU. Johnson meant to negotiate so as not to abandon the EU completely but to continue having some trade relations, as shown in (3) and (4):

> "Let us go for a deal that can heal this country, let's go for a deal that can heal this country and allow us all to express our legitimate desires for the deepest possible friendship and partnership with our neighbours" —19th October. PM statement in the House of Commons—.
> "A better deal that will maximise the opportunities of Brexit while allowing us to develop a new and exciting partnership with the rest of Europe based on free trade and mutual support" —24th July 2019. Boris Johnson's first speech as PM—.

Whereas the latter used the word *deal* to convince his voters that he would do his best to get things done, the former used the same word to attack what the PM promised. Related to this word and its sense, we can find instances of the lemma *abandon* in both corpora, as shown in figure 1.

The word *October* is closely linked to the Brexit process as it was the first date they proposed to leave the EU. For this reason, as can be seen in Figure 1, *October* is more frequent in Johnson's —10.7— than in Corbyn's —1.44—. The conservative leader was eager to fulfil his aim and he repeated the date, 31st October, in almost every speech we compiled.

Regarding Corbyn's corpus, two interesting words are worth analysing because they do not appear in Johnson's corpus, i.e., *tories* and *trump*. On the one hand, *tories* is the name commonly used to refer to conservative supporters. Corbyn uses this word to refer to their opponents, as shown in (5) and (6):

> "Who do you trust those who know the food industry inside out or Boris Johnson and the Tories?" —2nd September 2019. Jeremy Corbyn speech in Salford today—.
> "What a wonderful job Keir's done over the last three years, picking apart the Tories' shambolic handling of Brexit" —5th November 2019. Jeremy Corbyn's Brexit speech in Harlow—.

On the other, the word *Trump*, the US president's name, is also pronounced by Corbyn but

not by Johnson, as shown in (7) and (8). One of the most controversial policies that Johnson was will-ing to get once he gets Brexit done was to sign a trade deal with the US and its President, Donald Trump:

> "One that will see our industries run down, while the super rich get even richer from a Trump Deal Brexit" —2nd September 2019. Jeremy Corbyn speech in Salford today—.
>
> "He is doing nothing but seeking to divide, and risking this country's future for his own political gain and an America-First deal with President Trump" —3rd October 2019. Jeremy Corbyn responds to the Prime Minister's Brexit statement—.

As shown in figure 1, additionally, in Corbyn's corpus, there are words typically related to his ideology and policies, such as *jobs, nhs, climate* or *public*.

If we pay attention to the adjectives employed by each candidate, there are also signifi-cant differences. The most frequent adjectives used by Johnson are *great* —*great deal, great pro-jects, great United Kingdom*— and *possible* —*to make this possible, the best possible solution*—. Both of them referring to the Brexit process, as can be seen in (9) and (10):

> "I would prefer us to leave the EU with a deal. I would much prefer it. I believe that is still possible even at this late stage" —25th July 2019. PM statements on priorities for the government—.
>
> "And I want to stress that this is a great deal for our country, for the UK" —17th Oc-tober 2019. PM press conference at EU Council—.

However, the most frequent adjective in Corbyn is *real*, since *This is the real change* was the slogan for his political campaign. Additionally, the adjective *public* is also very frequent, as the la-bourist politician frequently calls on the public institutions in his speeches. We found collocations such as *public domain, public interest, public service workers, public health checks*, etc.

4.2. Sentiment Analysis comparison

As previously mentioned, to carry out the Sentiment Analysis, two different lexicons have been used: the NRC lexicon (Mohammad & Turney, 2013) and the BING lexicon (Hu & Liu, 2004). Therefore, results will be detailed according to each lexicon.

4.2.1. BING lexicon

According to the BING lexicon, both politicians' corpora contain more negative words than positive ones in their speeches. Boris Johnson's discourse is more negative at text level —2,893.5— than Corbyn's —1,573—, as can be seen in figure 2.
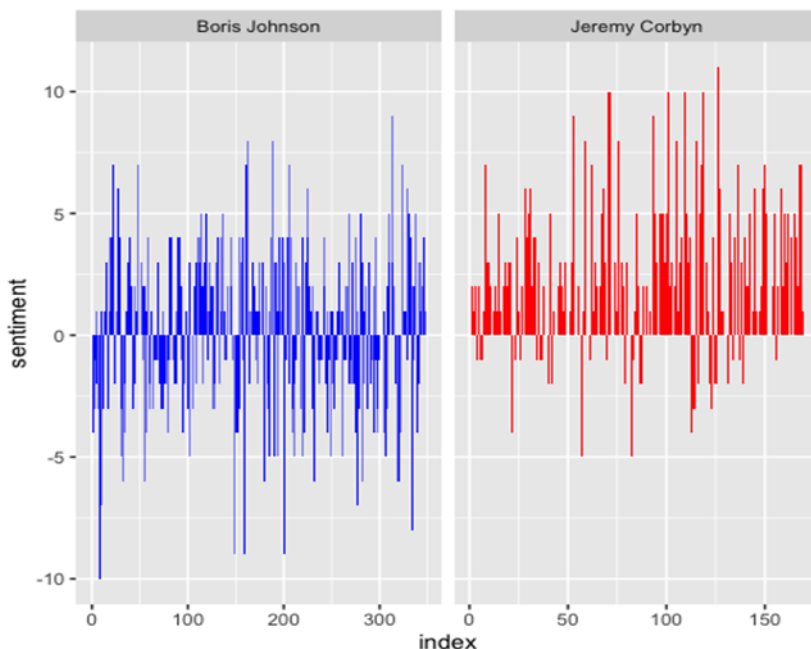
Figure 2. Comparison of Sentiment Analysis in Johnson's and Corbyn's speeches

Nevertheless, it is necessary to know which words carry that negative polarity to see why Johnson used more negative words than Corbyn. At word level, Figure 3 shows the top ten negative words of each politician. As can be seen, they are entirely different, except for the word *conservative,* which appears in both political discourse.
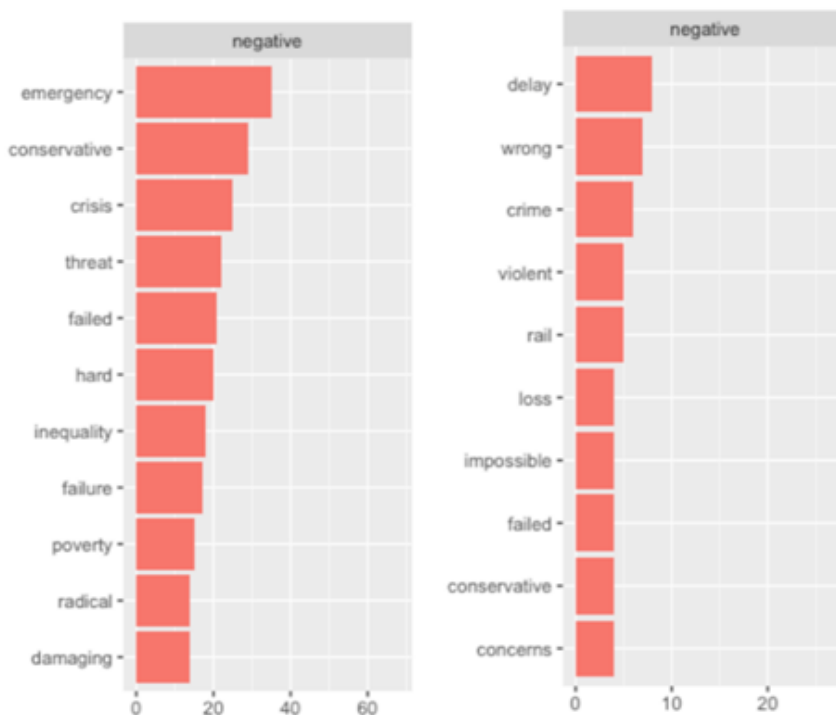

Figure 3. Top ten negative words for Boris (left) and Corbyn (right).

As shown in Figure 3, the BING lexicon classifies the word *conservative* as negative. According to the Cambridge dictionary, *conservative* means "not usually liking or trusting change, especially sudden change". However, in this context, it must be acknowledged that this word is not referring to this attitude but to the name of the Conservative Party, to which the PM belongs, and its ideology. Therefore, it should be classified as neutral.

The most frequent negative word in Johnson's corpus is *delay*. This word is really interesting in the Brexit context. As negotiations are taking so long, the Conservative leader uses this word to refer to the long period the Brexit process is taking. Its use can be seen in (11) and (12):

> "And this crucial mission there can no longer be any argument for further delay" —19th October. PM statement in the House of Commons—.
> "We are negotiating a deal and though I am confident of getting a deal we will leave on 31st October in all circumstances. There will be no further pointless delay" —3rd September 2019. PM statement on G7 Summit—.

Regarding Corbyn's corpus, *emergency* is the most frequent negative word. However, this word does not refer to Brexit but the climate change issue. In fact, the most frequent collocations, that is, words that frequently appear together, are *climate emergency, environment emergency* and *environmental emergency*, as shown in (13) and (14):

> "Mr Speaker, the UK should be using its position in the G7 to promote policies to tackle the climate emergency" —3rd September 2019. Jeremy Corbyn's emergency debate statement.
> "This election is our last chance to tackle the climate and environmental emergency" —18th November 2019. Jeremy Corbyn's speech at the Confederation of British Industry—.

In the same vein, the word *crisis* is also one of the most frequent and it also refers to climate change, with the collocation *climate crisis*. However, *crisis* is also used to get the audience's attention on the consequences that the Brexit process has brought with it. Corbyn refers to it as *the Brexit crisis*, as (15) and (16) illustrate, respectively. In fact, this is one of the most frequent collocation in his speeches:

> "And as we set out how our future economy will operate, we cannot ignore the most pressing issue of all: the climate crisis" —10th September 2019. Jeremy Corbyn speech to TUC Congress—.
> "The Brexit crisis needs to be resolved, but it must be done democratically" —5th November 2019. Jeremy Corbyn's Brexit speech in Harlow—.

Another interesting word that Figure 3 shows is the adjective *radical*. Corbyn uses this word to appeal for a new direction in politics. According to him, the British government needs a radical change when talking about the Brexit issue, as examples (17) and (18) show:

> "With a serious industrial strategy and a radical Labour government the economy can be a tool in our hands, rather than the master of our fate" —2nd September 2019. Jeremy Corbyn speech in Salford today—.
> "In this election, Labour is putting forward the most radical and far-reaching plan for real change in our lifetimes" —6th November 2019. Jeremy Corbyn's speech in Telford on leadership and what a Labour government will achieve—.

Besides, the word *inequality* is also among the most frequent negative words in Corbyn's political speeches, but it does not have any occurrence in Johnson's.

4.2.2. NRC lexicon

In terms of polarity, the NRC lexicon also classifies both corpora as negative, as many words bear a negative polarity, as shown in the previous section. One interesting word that the NRC lexicon, unlike the BING, assigned a negative polarity is *leave*. As expected, this word is more frequent in Johnson's speeches —7.12— than in Corbyn's ones —2.88—. Interestingly, although this word is assigned a negative polarity, it does only have such meaning in Corbyn's corpus. For John-

son, such word should be classified as positive or carrying the joy emotion, since it is their aim and allegedly the population's wish.

Other interesting words that are also negative but were not assigned such polarity by the BING are *force* or *wait* in Johnson's corpus. Both of them are closely related to the Brexit process and the problems they encountered in the EU. On the one hand, when it occurs as a verb, *force* refers to how the politician felt when talking about reaching a deal with the EU, as shown in (19) and (20):

> "That's what they want, to force us to beg for yet another pointless delay" —3rd September 2019. PM statement on G7 summit—.
> "would force me to go to Brussels and beg an extension" —3rd September 2019. PM statement on G7 summit—.

On the other hand, *wait* refers most of the times to the time that Brexit is taking to be done. Curiously, the PM used *people* as the subject of this verb and he also included himself in such group, as can be seen in (21) and (22):

> "But I know people can't wait and they want to see change faster" —27th July 2019. PM speech at Manchester Science and Industry Museum—.
> "We aren't going to wait 99 days because the British people have had enough of waiting" —24th July 2019. Boris Johnson's first speech as PM—.

In terms of emotion and at text level, unexpectedly, the predominant emotion of both corpora is that of trust, as can be seen in Figures 4 and 5. It is followed by the emotion of anticipation in both cases. However, the following emotion assigned to words is joy in Johnson's corpus but fear in Corbyn's.
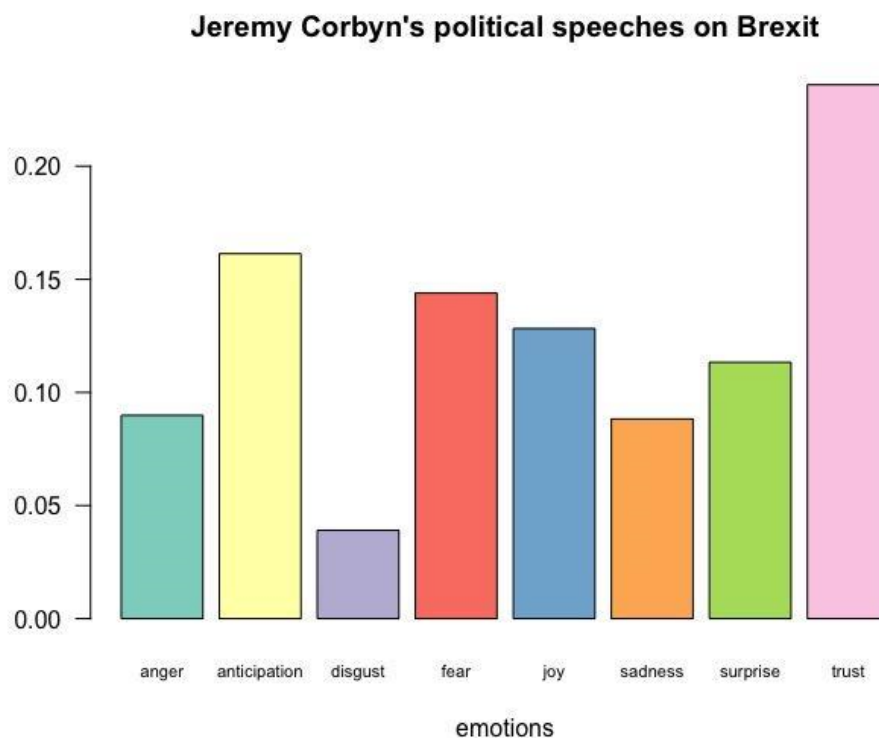

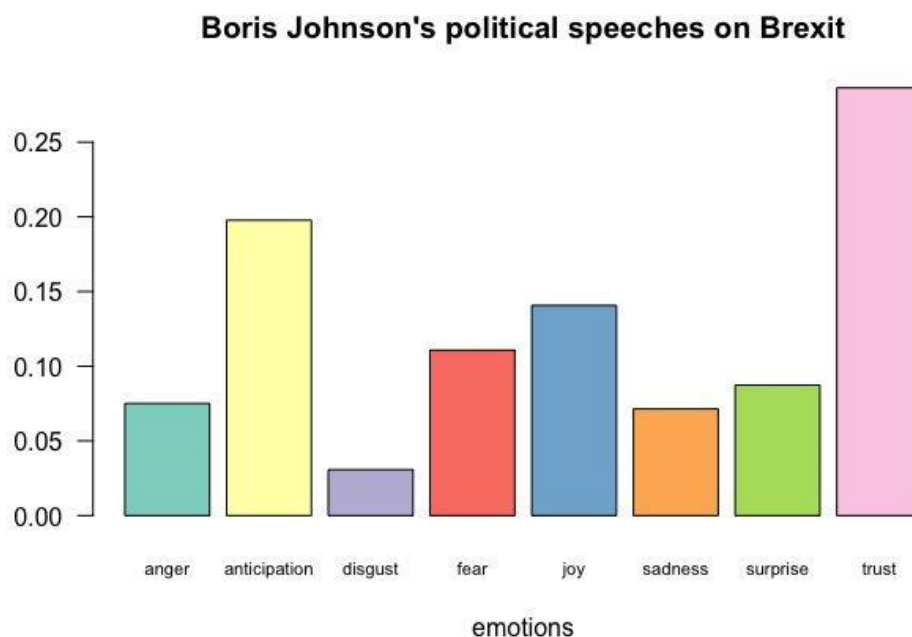Figure 4. Emotions in Jeremy Corbyn's political speeches on Brexit.

Figure 5. Emotions in Boris Johnson's political speeches on Brexit.

In both corpora, there are coinciding words that bear the emotion of trust, such as *deal, trade* or *parliament*, according to the NRC lexicon. However, this emotion is depicted by means of different words in each corpus. For example, concerning Johnson's corpus, some interesting words conveying trust are *nation* –8.31– and *hope* –8.31–. However, such words do not have a significant frequency in Corbyn's corpus –3.17 and 5.48, respectively–. Similarly, there are other words bearing the emotion of trust that are noteworthy in Corbyn's speeches.

As previously mentioned, *real* is very frequent in this corpus and it has been assigned the trust emotion. Additionally, *green* and *trust* itself occur in Corbyn's corpus more frequently than in Johnson's –11.2 vs 2.97 and 11.2 vs 4.75, respectively–. As previously shown, the word *green* refers to the environmental policies that Corbyn planned to implement if he had won. It has been demonstrated that he was really concerned about the climate change issue and everything related to it. The word *trust* is used mainly in two directions: questioning the conservative government's trust or concerning the population, as can be seen in (23) and (24):

> "Yes I believe leaders should have clear principles that people can trust". –6th November 2019. Jeremy Corbyn's speech in Telford on leadership and what a Labour government will achieve–.
> "Prime Minister, we can't trust you not to break the law because you've got form" –10th October 2019. Full text of Jeremy Corbyn's speech in Northampton–.

Regarding the other emotions, which are less prominent, some words are remarkable. Firstly, it is worth mentioning the word *ready,* which is very frequent in Johnson's speeches –11.2 vs 3.17 in Corbyn's–. The NRC lexicon assigned this word the anticipation emotion and somehow, it is used to anticipate that they can already leave the EU. Additionally, in Corbyn's speeches, we can observe that the word *change* has been assigned the fear emotion. However, as previously mentioned, *change* was part of the political campaign slogan, *The real change*. Thus, in Corbyn's corpus, such word should be associated with positive feelings rather than negative ones, such as fear.

5. DISCUSSION

As expected, both candidates' speeches show traces of each party ideology and this is conveyed through the word choice. The way language is used can reveal traits and intentions that politicians do not say overtly. According to Sorrnig, "the process of verbalising thoughts and transmitting ideas does not simply involve pure and unbiased statements" (1989, p. 95, as cited in Ekawati, 2019), but it is a constant declaration of purposes and aspirations. It is a task of Critical Discourse Analysis to reveal such intentions by interpreting the words used in the text. In politics, language is never neutral or unintentional.

Unexpectedly, *Brexit* and *EU* are among the most frequent words in both corpora and, in fact, the former was used as a selection criterion to filter the speeches. In Corbyn's corpus, other words related to his policies and ideology are also present, such as *NHS, workers* or *rights* or the continuous references to the climate change problem through a wide range of different terms. The electoral program of the Labour Party was divided into five blocks which are all of them reflected in the speeches. Firstly, under the title A green industrial revolution, they planned to tackle climate change paying attention to all the factors that contribute to it, such as Economy and energy, Transport, Environment and Animal welfare. Words such as the ones mentioned before: *green, climate change, climate crisis, environmental emergency* make reference to this point of the program. Secondly, there is a block devoted to public services, in which issues such as healthcare, security or education measures are included. As previously mentioned, words such as *NHS* or *workers* are very frequent in his corpus. Additionally, the adjective *public* was one of the most frequent. Thirdly, there is a package of measures to tackle poverty and inequality. In fact, these two words were among the most frequent negative words in Corbyn's corpus, as shown in Figure 3. As expected, there is also a block about Brexit and last but not least, there is one on internationalism.

However, in Johnson's corpus, no other topic than Brexit can be found. It seems that Boris Johnson has been so focused on getting Brexit done that he ignored other current affairs that are also important. In fact, in the Conservative Manifesto[7], the first intention we can see is "We will get Brexit done in January and unleash the potential of our whole country". Hence, it is not surprising that Brexit has been a priority during the whole political campaign. In the same vein, there are other words that reflect Johnson's ideology and intentions, such as *nation*, which is also used to make people feel part of it and support the Brexit movement.

Surprisingly, the word *workers* only has one occurrence and *rights* just five in the whole corpus. Such issues are fundamental in today's society and politicians should take care of them, especially if we are talking about the Prime Minister of a country. In relation to this, it is interesting to mention again that only 52% of the population voted in favour of Brexit, so almost half of the population does not want Brexit and is not probably interested in it. Thus, political speeches cannot be monopolised by the same topic since the population still has different and varied needs. The la-

---

[7] Available at: https://www.conservatives.com/our-plan.

bourist leader took advantage of this fact and in its Labour Manifesto[8], he claimed that the real change, that is, their policies are "for the many not the few".

As previously mentioned, one of the negative words used by Corbyn was *radical*. This word is generally interpreted with a negative connotation as it can be a synonym of *extreme*, which is related to *extremism*. However, in his speeches, Corbyn used this word in a positive way to indicate that there will be an enormous change in the country if people vote for him. The word *radical* is used to emphasise that the change will be in the opposite direction, as he did not want to leave the EU.

Regarding the overall sentiment of the texts, Johnson's contains more negative words than Corbyn's, according to the BING lexicon. However, paying attention to the analysis, Johnson used many words that generally have a negative polarity but not necessarily in this context as it happened with the word *leave*, which is one of the most frequent in the corpus.

Concerning the overall emotion assigned to both corpora, it came as a surprise that it was the one of trust. Generally speaking, not every politician gets to fulfil their promises during their term. For this reason, people do not usually believe all the things they claim they will do during their political campaign. In fact, one of the most common insults politicians frequently face is *liar* because of this reason. Politicians already know it, which may be why they try to use as many words bearing the trust emotion as they can to convince their electorate to vote for them and convey such emotion.

6. CONCLUSIONS

Politics consists of the communication between the candidate and their voters. Consequently, language is necessary so that politicians can transmit to their voters which measures they will take and communicate their intentions and promises during the political campaigns. Additionally, language is also used to reach agreements among the different parties, give explanations when needed, or even answer a different thing when asked a particular question. Thus, as known, language is closely linked to politics, and without language, politics could not occur. Needless to say, political discourse can reveal many of the candidate's intentions, but it can also show key information that influences the audience. Moreover, as this study demonstrates, the linguistic resources and keywords used during a political speech can have more or less impact on the voters.

As a result, politicians must carefully select the words to carry out all the actions mentioned above. For this reason, they are frequently helped by language experts so that the electorate could not be offended by the word choice. Thus, political speeches contain many words that reflect the political parties' ideology. In this research, we have found out that the most frequent words of each corpus clearly state the politicians' aims and the main measures of their political program.

Results show that Boris Johnson focused, almost exclusively, on the Brexit topic in his speeches, leaving aside other issues that are also happening in the country and the world and that probably affect the population. However, Jeremy Corbyn's discourse is more diverse and includes other

---

[8] Available at: https://labour.org.uk/wp-content/uploads/2019/11/Real-Change-Labour-Manifesto-2019.pdf.

issues he would have been focused on if he had won the election. Apart from the Brexit topic, other concerns such as the National Health Service or climate change are recurrent topics mentioned in his speeches. As is shown, the most frequent words of his political speeches remarkably correspond to the most important measures in his political program. Whereas Johnson's paid attention to just a little over half of the citizens, the 52% who voted in favour of Brexit, Corbyn's tried to cover most people's needs, as his slogan said, they made policies "for the many, not the few".

Indeed, the pandemic did not allow us to check what would have happened after Brexit day, once the primary purpose of the Conservative party was fulfilled since there were other current affairs to deal with. However, it would be interesting to see what Johnson will do once the pandemic will be finished, what measures or what policies he will implement, as he barely mentioned any of them in his speeches. Additionally, it would be necessary to check whether he continued governing for that 52% of people or change the government's direction to cope with the remaining citizens' needs.

BIBLIOGRAPHICAL REFERENCES

Baker, P. (2006). *Using corpora in discourse analysis*. A&C Black.

Desaulier, G. (2017). *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. Springer. https://doi.org/10.1007/978-3-319-64572-8

Ekawati, R. (2019). Power through linguistic modalities in Indonesian presidential speeches. *Discourse and Interaction, 12*(1), 5-28. https://doi.org/10.5817/DI2019-1-5

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language*. Routledge.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. *Discourse as social interaction*, 2, 258-284.

Fernández Melendres, C., & Moreno-Ortiz, A. (2020). Trump vs. Clinton: An Analysis of the US Presidential Debates of 2016. In D. Levey (Ed.), *Strategies and Analyses of Language and Communication in Multilingual and International Contexts* (pp. 141-152). Cambridge Scholars Publishing.

Fradejas Rueda, J. M. (2019). *Cuentapalabras. Estilometría y análisis de textos con R para filólogos*. http://www.aic.uva.es/cuentapalabras

Gerdes, J., & Stringam, B. B. (2008). Addressing researchers' quest for hospitality data: Mechanism for collecting data from web resources. *Tourism Analysis*, *13*(3), 309-315. https://doi.org/10.3727/108354208786094906

Hu, M., & Liu, B. (2004). Mining and summarising customer reviews. *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177. https://doi.org/10.1145/1014052.1014073

Jockers, M. L. (2015). *Syuzhet: Extract Sentiment and Plot Arcs from Text*. https://github.com/mjockers/syuzhet

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., & Suchomel, V. (2014). The

Sketch Engine: ten years on. *Lexicography*, *1*(1), 7-36. https://doi.org/10.1007/s40607-014-0009-9

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, *97*, 311-331. https://doi.org/10.1017/S0003055403000698

Lee, D. Y. W. (2008). Corpora and discourse analysis. In V. K. Bhatia, J. Flowerdew & R. H. Jones (Eds.) *Advances in Discourse Studies* (pp. 88-99). Routledge. https://doi.org/10.4324/9780203892299

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Mohammad, S. M. (2019). Word affect intensities. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 174-183. http://www.purl.org/net/saif.mohammad/research

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*(3), 436-465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

Moreno-Ortiz, A. (7-11 de septiembre de 2017). *Tecnolengua Lingmotif at EmoInt-2017: A lexicon-based approach*. Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Copenhagen, Denmark, 225-232.

Murillo Lanza, D. F. (2017). *Análisis de sentimientos de Cien años de soledad y El amor en los tiempos del cólera de Gabriel García Márquez* [Master dissertation, Universidad de Valladolid]. Repositorio documental. https://uvadoc.uva.es/handle/10324/25437

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1-135. https://doi.org/10.1561/1500000001

Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. *COLING/ACL 2006 - EMNLP 2006: 2006 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 327-335. https://doi.org/10.3115/1610075.1610122

Weiss, G., & Wodak, R. (2007). *Critical Discourse Analysis: Theory and Interdisciplinarity*. Palgrave Macmillan. https://doi.org/10.1057/9780230288423