Text-Mining Approach to Political Communication on Twitter

The Analysis of the Discourse of Spain's Principal Political Parties During the European Parliament Elections in 2019

Emilia Smolak Lozano[1] and Atsuho Nakayama[2]

[1] *Department of Audiovisual Communication and Advertising, Faculty of Communication Sciences, University of Malaga, Malaga, Spain;* [2] *Department of Business Administration, Graduate School of Management, University of Tokyo, Tokyo, Japan*

[1] *https://orcid.org/0000-0001-8193-8786;* [2] *https://orcid.org/0000-0002-9220-5059*

Abstract

Twitter has become a powerful tool of political communication, that now plays a significant role during elections, especially in countries such as Spain, where use of digital media is extended widely throughout society. Digital democracy is based to a significant extent on the quality of public discourse and persuasion implemented in the digital messages contained in tweets. Text mining methods applied to tweets during the 2019 European elections made it possible to examine content, frequently used keywords and expressions, sentiment and tone of the political discourse of the main Spanish political parties. The objective of the analysis is to determine the scope and thematic focus of the political discourse on Twitter and make an inter-party comparison. The results reveal that Spanish politics were a much bigger focus than the European perspective and the social outlook pursued by the left wing turned out to be more visible than other proposals. Fragmentised discourse in the case of the populist parties focused on concrete problems to be resolved, whereas the main approach of Twitter politics was the fight against right-wing rivals. It is possible to conclude rather low maturity in terms of democratic public discourse with the high persuasive components integrated within tweets and a self-appraising attitude.

Keywords

social media, Twitter, text mining, discourse analysis, political communication, political public relations

Introduction

Democracy in general is the political and institutional framework for political communication. The development of new forms of communication based on online technologies has established the foundations for cyberdemocracy within the wider background of the information and knowledge society. As such, the phenomenon of electronic democracy has brought new forms of political communication to the digital ecosystems, described as the digital infrastructure in contemporary politics or as hybrid politics (combined with traditional mass media). The internet is considered to be an effective channel of political communication, especially when aimed at reaching key digital audiences, disseminated across digital platforms and ecosystems.

Political organisation-public relationships together with political public relations role in political communication constitute the useful framework to approach the use of social media in the strategies of political campaigns (Browning & Sweetser, 2020; Froehlich & Rüdiger, 2006; Hong, 2013). Applying online political public relations to campaigns (Painter, 2015) can contribute to agenda setting (Kiousis & Shields, 2008) and message framing or help increase the participation of young voters (Dong & Ji, 2018). In general, public relationship (PR) practice seems to be essential in digital political communication (Erzikova & Bowen, 2019) and use of social media in public oriented strategies can be effective, as has been shown in some research (Allagui & Breslow, 2016). As such, online political public relations, or in a more general manner, social media must now work in synergy with traditional media, in order to maximise digital influence in hybrid political communication (Rune & Enjolras, 2016).

Social Media in Political Campaigns

Social media not only offers a means of achieving more dialogical and bidirectional communication (Plowman et al., 2015) but can also become useful tools for political campaigns (Housholder & LaMarre, 2015). When applied for the purposes of political branding (disseminating the desired image of political parties to certain segments of society), they can position themselves effectively among the target voter groups (Baines, 1999; Cwalina & Falkowski, 2015). In these terms, Obama's campaign during the 2008 election brought a revolution to political communication (Macnamara, 2012) similar to Trump's political campaign, where 44% of budget was allocated to digital media. As such, social media has become the commonly-used strategy in modern political campaigns but with limited investment to date (7% of all spending in political campaigns in the USA in 2018), however with promising forecasts for increase from 2018 onwards, especially in U.S. politics. Online channels play a significant role in political discourses in the networked society (Campbell & Kwak, 2011) whereas candidates' rhetoric is a key element in political strategies (Sisco et al., 2017).

Macnamara (2012) has analysed implementation of web 2.0 in political elections and identifies the need to work more on stimulating and seeding online political discussions, elaboration of tools to improve development of the arguments and correct implementation and analysis of user's data and big data. Practitioners need to learn new skills to be effective in social media, including informal "conversational" styles of writing online,

new techniques of media relations (e.g., bloggers do not come to news conferences and most don't accept press releases) and online community management techniques (Macnamara, 2012).

Nevertheless, social media is used in political campaigns not only to create and disseminate the image of political candidates or parties (Kelm et al., 2019; Rune & Enjolras, 2016) but also for framing the message and discourse in social networks, especially on Twitter or Facebook (Stetka et al., 2019; Vesnic-Alujevic, 2012). This application as the tool of agenda setting (Lee & Xu, 2018) takes place due to its discursive power and media effects (Bulovsky, 2019; Sahly et al., 2019).

As such, Twitter and its public sphere provides a scenario that makes it possible to observe the extension and effects of the stories and discourses within political communication (Guo et al., 2019). Fontecilla Camps (1988) identifies four main elements of political discourse; as a result, electoral discourse will be mainly focused on persuasion, ideology, and pro-receptor. For instance, analysis of digital political advertising indicates three main issues that frame the federal political discourse disseminated by ads on Facebook or Google in the USA in 2018: health, taxes, and Medicare. Roper and Hurst (2019) claim that correct PR practice can be helpful to establish a useful and dialogue-based discourse — "political talk" — even in the case of complicated issues. It is therefore important to examine the political discourse as shaped during political campaigns on online media (Wojcieszak & Mutz, 2009): their contents, topics, key issues, framing, and agenda in order to introduce any improvements strategically. The textual data that are generated during the campaigns constitute the useful field for research, for content and textual analysis based on text mining and a big data approach in modern PR management strategies. It can therefore be a helpful tool to examine, as Deligiaouri (2018) points out, the phenomenon of discursive construction of ideology and post-truth narratives in the contemporary political communication that primarily takes place in the digital ecosystem (Speakman, 2015).

Similar studies on the implementation of social media in campaigns, in particular social networks, have studied its role as a public relations strategy (Frame & Brachotte, 2015; LaMarre & Suzuki-Lambrecht, 2013). Enli and Skogerbø (2013) examined Facebook and Twitter as the arenas of political communication. Xiong et al. (2019) have examined Twitter by means of semantic analysis of message framing, using hashtags and topics. As far as text analysis methods are concerned, Xu and Xiong (2020) apply text analysis to approach a polarised discourse on Twitter regarding social

and political issues preceded by Adams' and McCorkindale's (2013) content analysis of Twitter usage in presidential campaigns or Pressgrove's and Kim's (2018) content analysis of the 2016 elections. Studies of the use of tweets and retweets in campaigns were performed by Lee and Lim (2016) or Lee and Xu (2018). Rune and Enjolras (2016) analysed campaigning styles on Twitter, differentiating the personalised strategies of candidates versus parties' reputation and narratives. Casero-Ripollés et al. (2017) examined the populist discourse of the Unidos Podemos party on Twitter during the 2016 Spanish elections. This is one of the few studies focused on Twitter discourse analysis in the Spanish context. In these terms, Wang (2016) points out the new principles for political agenda and political discourse analysis, focusing on language processing and computational linguistics as key techniques for critical studies of political narrative in social media.

Text Mining and Analysis of Political Discourse

Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools (e.g., Feldman & Sanger, 2006). Text mining extracts useful and important information from texts by identifying and exploring significant rules. Any text from a paper, essay, book, newsletter, email, or post of social networking service may be used, in function of the researcher's goals. These document collections are not the formalised database records but the unstructured textual data. However, we must find interesting patterns from them. Feldman and Sanger (2006) pointed out that text mining is very similar to data mining and derives much of its inspiration and direction from seminal research on data mining. It is therefore not surprising to find that text mining and data mining systems have many high-level architectural similarities. Much of their pre-processing focus falls on scrubbing and normalising data and creating extensive numbers of table joins, because data mining assumes that the data has already been stored in a structured format. In contrast, for text mining systems, pre-processing operations centre on the identification and extraction of representative features for natural language documents. These pre-processing operations are responsible for transforming unstructured data stored in document collections into a more explicitly structured intermediate format, which is a concern that is not relevant for most data mining systems. Netzer et al. (2012) introduced the idea that business fields have been expanding the opportunities of use of text mining approaches to collect competitive intelligence, to analyse the wealth of information that consumers are posting online, and to

analyse the infinite stream of financial report data, to search for patterns or irregularities. The expansion of these opportunities is based on the fact that the availability of digital text data is increasing. Similar approaches can be found in the studies of Nicholls and Bright (2019) in which content analysis and cluster analysis were applied, in the qualitative analysis (critical thematic analysis) of Lawless and Chen (2019). Further foundations regarding the political communication research via textual analysis can be established by critical discourse analysis (Reynolds, 2019) or textual and contextual analysis (Trimithiotis, 2018). However, the present study also focuses on using a text-mining approach by applying the computational techniques of language processing in order to assess the sentiment and contents from tweets related to populism discourse in social media in the European elections in 2019.

The Origins and Outline of the Study

Inspired by the comparative study by Jansen et al. (2019) on the European Parliament elections, the present chapter examines the political agenda and discourse performed by Spanish political parties in social media during the European Parliament elections in Spain in 2019, using text mining techniques applied to tweets published by the political parties during the campaign. As such, the study attempts to evaluate the degree of populist discourse in European modern politics: from Euroscepticism to pro-European positions in Spain, given the increasing popularity of extreme or populist parties in Spain, on both the left and right wing. While the media-based or journalistic discourse on politics is found in scientific research (Reynolds, 2019; Tameryan et al., 2018, to name the most recent examples), the social media narrative of political parties has yet to be explored. Echeverría (2017) notices that media frames and discourse during elections focuses on the infotainment dimension rather than the proposals and themes of the candidates' campaigns. Therefore, it is extremely important to examine the sentiments and proposals present in political discourse directed to the European level of politics.

Following the attempts made by Deligiaouri (2018), the research therefore employs advanced text mining techniques to analyse the discursive construction of the narrative in contemporary politics on Twitter by political parties running for European Parliament. In this way, multi-dimensional and cluster analysis techniques will make it possible to identify the main frames, sentiments, themes, and focuses that dominated the political discourse in

the Spanish political arena during the elections, as a reflection of the interests existing within the political parties. Similar research on political discourse during the European elections was performed by Trimithiotis (2018), using the textual and contextual perspective, but focused on Europe as a main topic. This analysis aims to determine whether the narrative of different political parties in Spain presents any similarities or variations and to what extent. Moreover, it will help discover the degree of the communication flow in political discourse in Spain in relation to populism, Euroscepticism, polarisation, and progressive narration within the digital political discourse in Spain. Using Twitter as the main tool for discursive expression and the construction of dialogue (Watkins, 2017) between political actors and citizens (Gálvez-Rodríguez et al., 2018) in cyber-politics can possibly help overcome the bias between political representation and public opinion (Druckman, 2014). For this reason, it is first important to determine the sentiment and linguistic framework of the discourse.

Methodology

The study focuses on the European Parliament elections held between May 23–26, 2019, in all European Union (EU) member states (27 countries). The campaign period in Spain was from May 9–23, 2019, together with the election process. The campaign focused on political parties and their programmes. The importance of political campaigns during the EU Parliament elections is supported by the turnout data. According to the European Parliament (2019) on average, the turnout was 51% in the EU and 61% on average in Spain. Spain indicates widely pro-European attitudes (75% of population), but with an increasingly growing populist discourse. As such, we have obtained a vast representation of dominant perceptions and attitudes in political communication in Spain in relation to the EU, represented by parties and their voters with more recent attention to populism and its discourse as a dominating ideology.

The study analysed the tweets published by the selected political parties during the EU Parliament election campaign (from May 9–23, 2019). The tweets were downloaded from Twitter using a special programme and the most relevant terms were then sub-weighted after using specially designed algorithms to remove the punctuation symbols, signs and prepositions for each language. The study's main objective is two-fold:

1. to determine the scope and thematic focus of the political discourse on Twitter during the EU elections in Spain;

2. to perform a comparative analysis of the discourse of the main Spanish parties contained in the tweets published during the EU campaign in 2019.

The Twitter database consisted of 2,052 tweets in the case of Spain. The numbers per party were distributed as follows (Table 1).

Table 1 Tweets data distribution.

1 ciudadanos_automony_may.csv	591 tweets
2 psoe_automony_may.csv	587 tweets
3 populares_automony_may.csv	404 tweets
4 ahorapodemos_automony_may.csv	294 tweets
5 vox_automony_may.csv	176 tweets
Total	**2,052 tweets**

In this study, we perform analysis of post tagging, lemmatisation, and co-occurrences. In this vignette, we will show some basic frequency statistics that can be extracted once we have annotated our text. We use the UDPipe R package, which provides language-agnostic tokenisation, tagging, lemmatisation, and dependency parsing of raw text, which is an essential part in natural language processing. Furthermore, the logic accounts for all languages and is language-agnostic[1]. Co-occurrences make it possible to see how words are used either in the same sentence or next to each other. The UDPipe package makes it possible to create co-occurrence graphs using the relevant parts of speech tags. We look at how many times nouns and adjectives are implemented in the same sentence. We visualised the result by using the ggraph R package, that can visualise the word network[2]. Once we get these co-occurrences, we can perform the same plotting using the ggraph R package.

Firstly, we started by annotating some text in English. The annotated data frame can then be used for basic text analytics. A data frame is a table or a two-dimensional array-like structure, in which each column contains values of one variable, such as numerical vectors and character vectors, and each row contains one set of values from each column. Although it looks like a two-dimensional array, like a matrix, it differs from a matrix in that each

---

[1] For more details, please see the reference manual of UDPipe (Wijffels, 2021) or site of basic analytical use cases of UDPipe (BNOSAC, n.d.).

[2] For more details, please see the reference manual of ggraph (Pedersen, 2021).

row and column of the data frame must have a label and can be manipulated by labels.

The resulting data frame is one row per "doc_id" and "term_id", containing all the tokens in the data, the lemma, the part of speech tags, the morphological features, the dependency relationship along the tokens, and the location where the token is found in the original text. A field called "upos", which is the universal parts of speech tag, and a field called "lemma" which is the root form of each token in the text, give us a broad range of analytical possibilities.

Frequency statistics of words brings good results, but we need to identify words that make sense in combination with other words. We must therefore confirm keywords that are a combination of words. In the UDPipe package, we can identify keywords in the text by following three methods: rapid automatic keyword extraction (RAKE; Rose et al., 2010), collocation ordering using pointwise mutual information (PMI; Church & Hanks, 1990), and parts of speech phrase sequence detection. Therefore, we used these three methods to identify keywords in the text. RAKE algorithm is one of the most popular (unsupervised) algorithms of machine learning for extracting keywords during information retrieval. It is a domain independent keyword extraction algorithm, which tries to determine key phrases in a body of text by analysing the frequency of word appearance and its co-occurrence with other words in the text.

Collocations are a sequence of words or terms that co-occur more often than would be expected by chance. Common collocations are adjectives + nouns, nouns followed by nouns, verbs and nouns, adverbs and adjectives, verbs and prepositional phrases or verbs and adverbs. By computing PMI which are indicators of how likely two terms are collocated compared to being independent, we can extract relevant collocations. The PMI of a pair of outcomes "x" and "y" belonging to variables x and y quantifies the discrepancy between the probability of their coincidence given their joint distribution and their individual distributions, assuming independence. The PMI formula is:

$$PMI = \log \frac{p(x, y)}{p(x)p(y)}.$$

In this way, we analysed the content, the most frequently used keywords and expressions, sentiment and tone of the 2,052 tweets published by the principal Spanish parties (Partido Popular [PP], Partido Socialista Obrero Español [PSOE], Ciudadanos, Vox, Unidos Podemos) during the EU Parliament election campaign in May 2019.

The Results

The analysis is divided into two parts. Firstly, the general political discourse in Spain will be examined. Secondly, there will be a comparative study of the discourse of political parties in Spain during the European elections, primarily using keywords and co-occurrences techniques.

Spanish Political Discourse in the European Elections

In most languages, nouns are the most common types of words, next to verbs and these are the most relevant for analytical purposes, next to adjectives and proper nouns. Figure 1 shows the frequency of occurrence of universal parts of speech (UPOS). The nouns are the most frequent, next to adpositions.
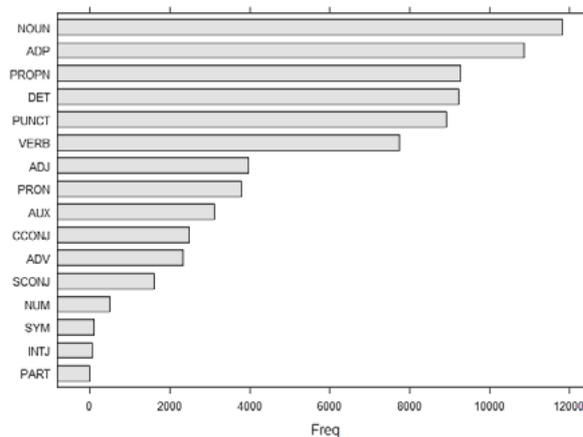

Figure 1 The frequency of occurrence of UPOS in Spanish tweets.

We can therefore confirm that the most common words were nouns, be-
cause we obtained the text annotated with parts of speech. The language
is oriented by the objects and concepts, rather than actions or expression
of will to take an action.

Figure 1 shows the top 20 occurring nouns. It indicates that these nouns are
frequently used in tweets.
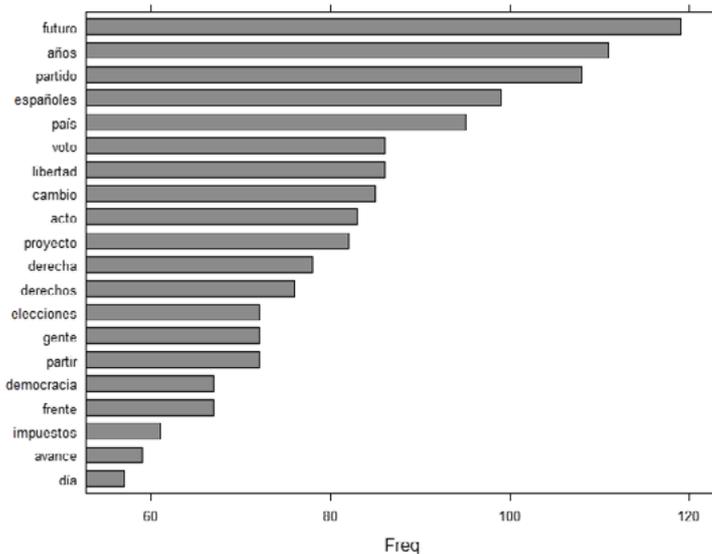
Figure 2 The top 20 occurring nouns.

The most frequently used words are "future", "years", "party", "Spanish citi-
zens", and "country", followed by "vote", "freedom", "change", "act", and "proj-
ect". "Law", "rights", and "democracy" are far less present in the discourse.
They were used fewer than 80 times. The frequency analysis does not indi-
cate any main topics that guide the discourse or does not focus on any ur-
gent issues that are important for citizens. Instead, the frequency of words
indicates the persuasive tone of the messages and its framework points to
general political discussion: party, country, future, and Spanish citizens.

Human speech often tends to exaggerate the object with an adjective, and therefore it is necessary to look at the most frequently occurring adjectives. Figure 3 shows the top 20 occurring adjectives. It indicates that these adjectives are frequently used in tweets.
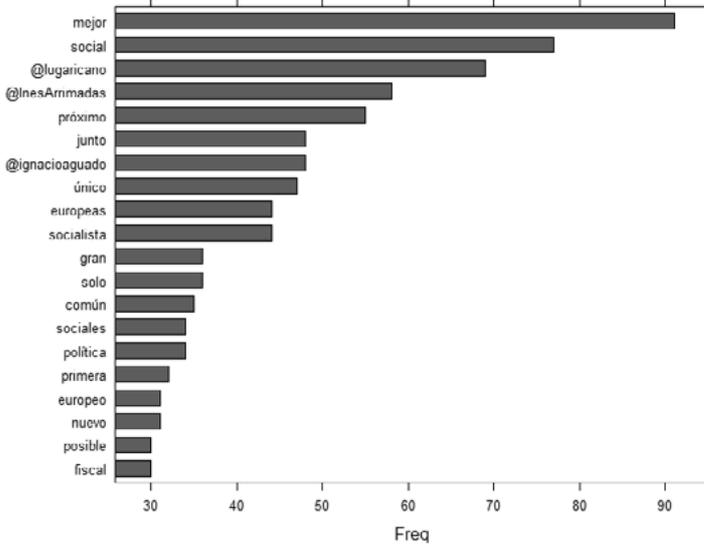

Figure 3 Top 20 occurring adjectives.

Spanish language applies adjectives quite frequently. The most frequently used adjectives are "better" or "social". The words "next", "together" (jointly), "unique", and "European" are also popular. This indicates the orientation towards progressive ideas denoting a better and more social future, in which the community perspective dominates.

We can reveal the nature of the tweeting by checking the usage of verbs. The usage of verbs indicates whether there is any sign of optimism or simply infuse pessimism. Figure 4 shows the top 20 occurring verbs. It indicates that these adjectives are frequently used in tweets.
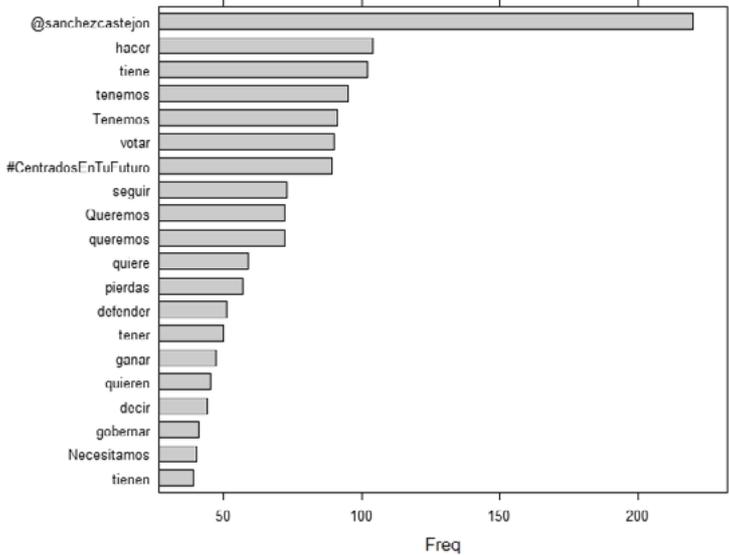
Figure 4 The top 20 occurring verbs.

As we can see, similar to Figure 3, the programme is not yet ideal in rec-
ognising the parts of the speech in different linguistic context within the
tweets. The Twitter replies beginning with @ and including a person to
whom the message is directed, as can be observed on both graphs, is pro-
cessed by the linguistic computer techniques as verbs or adjectives. It also
detects hashtags as verbs.

"To do" and "to have" are the two most frequently used verbs in this case.
This is followed by "to have to", followed by "to vote", "to follow", and "to
want". The declarative aspect of the language framework of the discourse
is dominant however, pointing to some action or offer/things in common
as well as the sense of obligation imposed on the social group. Clearly, the
voting intention is marked, as well.

Figure 5 shows the top 20 key phrases identified by RAKE. The key phrases
were extracted on the basis of the condition that the frequency was higher
than three. It indicates that these key phrases are frequently used in the
body of tweets.
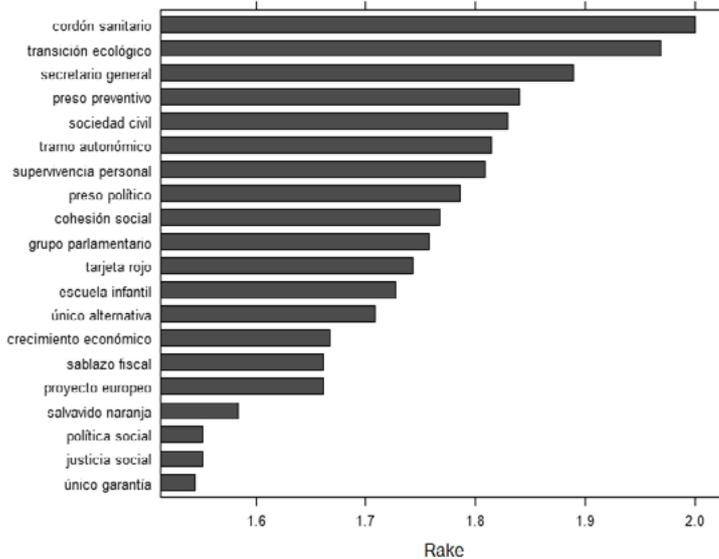
Figure 5 The top 20 key phrases identified by rapid automatic keyword extraction.

Key phrases detected by linguistic computational techniques in Spain reflects the political debate between the political parties. The first key phrase to appear is "sanitary cordon" as the reference to isolate the populist parties (in this case the right-wing Vox party). Furthermore, we can detect the key phrase of the political programme that is "ecological transition". The discourse mentions the Catalan situation ("preventive prisoners" and "political prisoners") as well as issues related to the democratic order in Spain and its structure: "civil society", "autonomies", "social cohesion". The following expressions are related to the political tension present in Spain in 2019. First, the clear party-centred discourse with the elevated use of the phrase "parliamentary group". Secondly, the expression "red card" — a famous gesture by the president of Spanish government, Pedro Sanchez, towards three right-wing parties — to encourage left-wing voters to prevent the right wing from winning the election and form a coalition due to the danger to social rights. A clear trend is also observed. The key issues of the political programme that shall form the foundation of the political discourse in elections are those that are far less frequently used — "economic growth", "taxes reform", "European project", "social policy", and "social justice". The discourse

is centred on the political issues, conflict and rivalry between the Spanish political parties running for European Parliament without barely denoting the most important topics of interest for citizens.

Figure 6 shows the top 20 keywords identified by PMI collocations. The key phrases were extracted under the condition that the frequency was greater than three. We could extract relevant collocations. These keywords show the sequence of words or terms that co-occur more often than would be expected by chance.
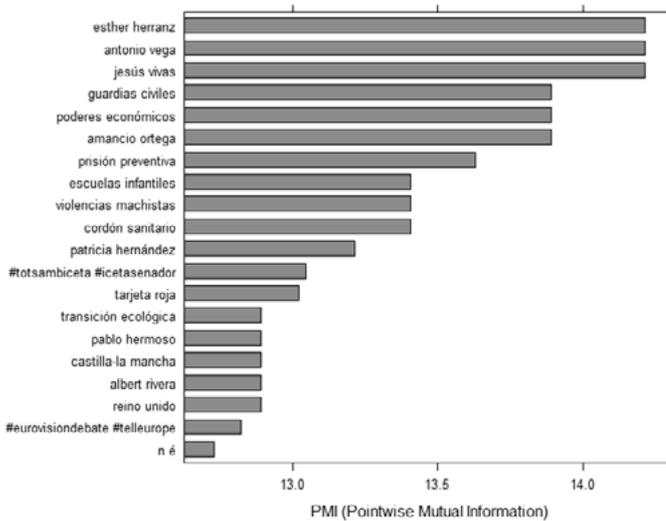
Figure 6 The top 20 keywords identified by PMI collocations.

The keywords detected by applying PMI collocation mainly indicate the names of politicians (the former European parliament member Esther Herranz), artists (Antonio Vega, 10th anniversary of the death of this musician, who wrote the song *Giant's Battle*, used as an election metaphor). The keywords that appear include the military police corps (the Guardia Civil), economic powers, Amancio Ortega (Spain's richest entrepreneur, the founder of Zara), preventive prison (reform of the penal code in Spain), kindergarten (social project to help families), gender violence, and sanitary cordon (towards populist parties). The political discourse is personalised

and oriented towards certain politicians or celebrities. Its fragmentation reveals multiple issues that are being talked about with a clear degree of polarisation and personalisation (economic powers and Amancio Ortega) versus progressive social reforms. Some urgent issues of Spanish politics are recorded (penal code reform). There is no clear reference to European issues or problems — except for the United Kingdom which is almost the least mentioned item.

Probably, in many languages, a simple noun and a verb form a phrase. We understand the context of the sentence by a phrase such as "go voting" that consists of the verb "to go" and the noun "voting". We can highlight top phrases by reverse engineering using the tweets data. Figure 7 shows the top 20 keywords of simple noun phrases. Figure 8 shows the top 20 keywords of simple verb phrases. The key phrases were extracted under the condition that the *n*-gram was greater than one and the frequency was greater than three. The *n*-gram shows a contiguous sequence of *n* items from a given text in natural language processing.
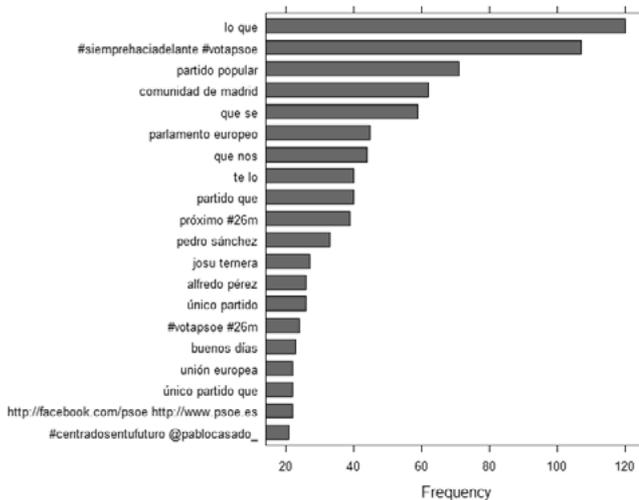
Figure 7 The top 20 keywords of simple noun phrases.

Firstly, it is necessary to adjust the technique for each language given its specific nature, as can be observed in the above graphic. Otherwise, the programme will detect semantic structures that, for example, are used in Spanish to put together sentences. Given these constraints, the analysis will ignore those structures and focus only on the properly detected items. First, the more popular noun phrases were the hashtags of PSOE´s campaign — "#foreverforward" and "#votePSOE": main campaign's message combined with persuasive sentence. Then, we observe the phrases that denominate the right-wing "Partido Popular", the "Madrid Region", "European Parliament", "political party", and the names of the political leaders. Apart from persuasive messages regarding the socialist party, most of the noun structures are the institutional actors of the political campaign — the people responsible for the political messages — who point to themselves.
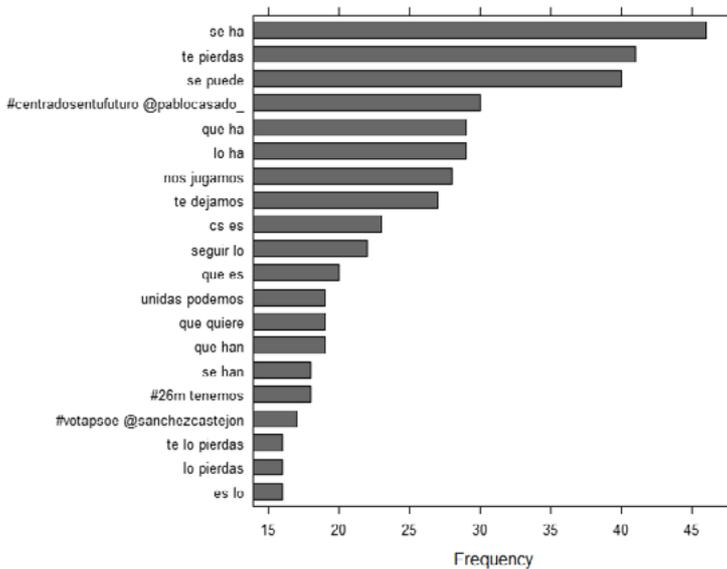
Figure 8 The top 20 keywords of simple verb phrases.

Figure 9 shows the top 20 co-occurrences within sentence. The edge (pass or link) width indicates the degree of co-occurrence between words. For

example, the results show the use of the following as the main framework of the discourse among the Spanish tweets: "European elections", "social justice", "the only party", and "future progress". "European elections" was the phrase most frequently used. We can also see that "extreme right", "autonomous region", "populism, nationalism", "candidate", and "debate" were other key phrases. Phrases such as "climate change", "coexistence" also appear, but to lesser degree. The co-occurrences most frequently used in the discourse is "populism" together with "nationalism" in different structures. As a result, these two concepts become equal to each other.

Figure 9 Top 20 co-occurrences within sentence.

If we are interested in visualising which words follow one another, this can be done by calculating word co-occurrences of a specific parts of speech type, which follow one another where we can specify how far the researcher wants to look, in terms of "following one another". Figure 10 shows the top 20 nouns and adjectives which follow one another. In this analysis, we look how many times the nouns and adjectives are used in the same sentence. We set skipgram at one. It means looking to the next word and the word after that. Edge (pass or link) width indicates the degree of co-occurrence of nouns and adjectives which follow one another.

Figure 10 Top 20 nouns and adjectives that follow one another.

Keyword correlations indicate how terms are placed together in the same document/sentence. While co-occurrences focus on frequency, correlation measures between two terms can also be high, even if two terms occur only a small number of times, but always appear together. Figure 11 shows the top 20 correlations between words within each sentence. We used the ggraph R package to get the same plotting as above. In this analysis, we reveal how nouns and adjectives are correlated within each sentence of a document. The edge (pass or link) width indicates the degree of correlation between words. The words most commonly following others are "social" related to "justice" or "rights", "European" referring to elections, "climate" in regard to "change", "extreme" for the right-wing parties or "autonomous" for regions or government (the basic political structure in Spain).
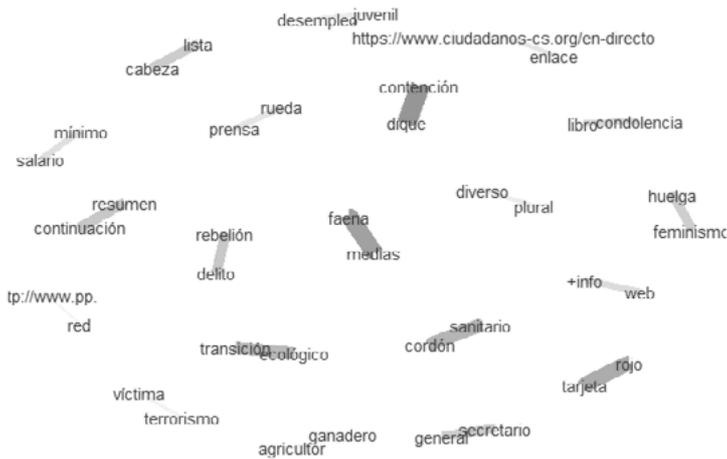
Figure 11 Top 20 correlation between words within a sentence.

The results show the use of words regarding the political elections "European elections", "social justice", "autonomous region", and "the only party". In particular, the other the most frequently used expressions are the following: "climate change", "the best team", and "extreme right wing". These results confirm the most popular noun-adjective combination, revealing the real degree of appearance. Thanks to this technique, we can also see that more expressions are included within the discourse: public service, terrorism victims, parliamentary groups, autonomic government, and coherent vote. We can then understand that some words are the key phrase. For instance, these phrases are "European project and European elections", "social rights and social justice". However, the discourse is mainly oriented to political actors and processes themselves, pointing out the winners, persuading votes and lacks political and social topics to guide and frame the political debate. Only "social issues" seem to be firmly stated across the tweets, as well as the "extreme right". In this sense, there is a clear indication for polarisation of the political narrative.

Spanish Political Parties´ Discourses During the European Elections

In the second stage of the study, we compared the discourse of the main Spanish parties running for EU parliament: PSOE (socialists), PP (Christian democratic), Ciudadanos (liberal), Unidos Podemos (extreme-leftists), and finally Vox (extreme-right).

Vox
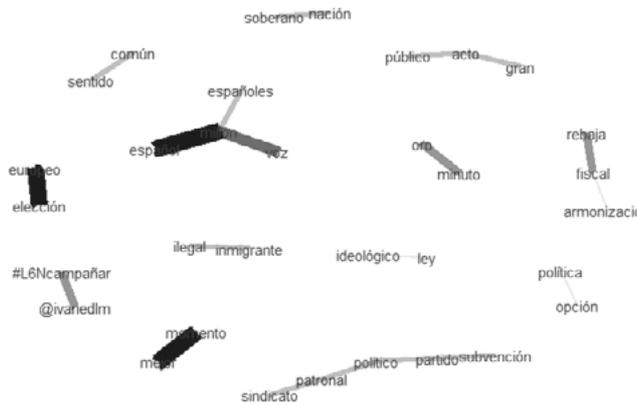
Figure 12 Nouns and adjectives combination in Vox' discourse.

Analysis of noun and adjectives combinations and their frequency (words following one another) in the case of Vox reveals the main axes of the discourse: the voice of the Spanish people, common sense, sovereignty, Vox as the only political option and the best moment for change (European elections), tax reform (lower taxes). These are main persuasive messages which suggest Vox as the political option for reasonable citizens to protect the sovereignty of the country in the EU and suggesting economic reforms. Part of the discourse is directed towards its rivals, calling them the parties that are subsidised by public funds, against trade unions, pointing out the law constructed on the ideological foundations or illegal emigrants.
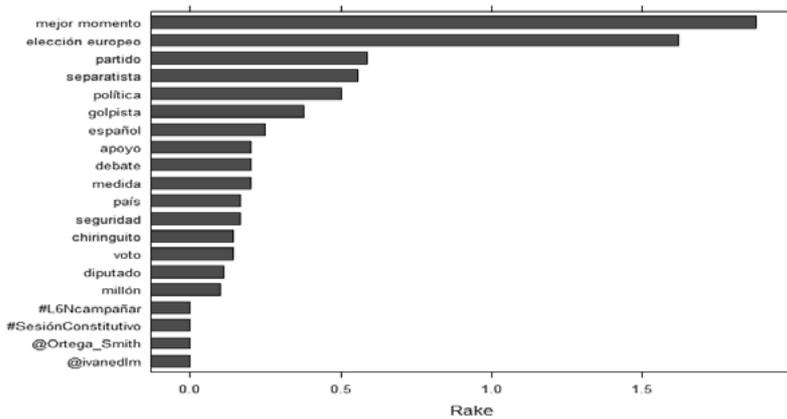
Figure 13 Top keywords in Vox's Twitter discourse.

The keywords analysis of RAKE demonstrates that Vox considers that the European elections is the best moment of change and talks primarily about the parties as the political actors. The fourth most frequent word is "separatist" politics, which is the main concern for right-wing populists in terms of the country's integrity. The frequently used adjective "Spanish" reflects nationalism. Moreover, it talks about security of the country and its citizens and the legally-dubious companies that have been created by the adversaries of the Vox party.

The analysis of co-occurrences within the sentence of nouns and adjectives emphasises the institutional orientation of discourse, focused on political parties, candidates and the electoral process ("debate", "party", "candidate", "European elections", "votes"). Nevertheless, the main persuasive message was shaped by the common denominator of reasonable thinking and sense of community that can fight back against impositions: common front, sovereignty or common sense.

Keywords correlations between nouns and adjectives additionally reveals the most frequently appearing topics, such as Vox as the party that leads the country, nation, the May elections, the voice of a million Spanish citizens by choosing Vox, nation (part of the nationalistic and patriotic narrative), and crisis (stressing the current political and economic situation in Spain). It also mentions the party's leader, Ortega Smith.

In summary, Vox points out the country's current political and systemic problems, presenting itself as the rescue for people who care for their country and think reasonably about the possible solutions. In these terms, the EU elections and the choice of Vox is presented as the hope for Spain to solve its sovereignty problems.

Unidos Podemos

Unidos Podemos' political discourse, as shaped by the party during the EU elections (analysis of words following one another) is focused mainly on social issues and problems: social justice and social rights. It presents itself in a persuasive manner as part of the progressive coalition: the only choice for the majority that cares for social equality, progress, and change. Hence, the campaign is the opportunity for change that lies in the hands of voters (inspired by Obama's hashtag slogan: #YouCanChangeEverything). Apart from the institutional dimension (institution, European, elections) the message includes social issues, such as gender violence, climate change, public service, and (public) healthcare. This is quite different from the general political discourse in the Spanish EU elections of all parties altogether that lacked such mentions. The narrative is clearly pointed against the extreme or populistic right (*extrema derecha*) as the main rival to Podemos, part of the discourse is directed against this party. Keywords identified by RAKE confirmed that the main recipient of the discourse of the "social majority" (those who cares for social progress and right) was the main voters of Podemos. "Extreme right-wing adversary" is the fifth most frequently mentioned keyword. As a result, Vox constitutes a significant part of Podemos' narrative. The main political issues are social justice, climate change, and public healthcare, whereas gender violence and public service are less frequently used in Twitter messages.

Furthermore, keywords correlations indicates more dimensions of the Unidos Podemos' main electoral narrative: Unidos Podemos is the unique horizon for politics as the party of common force and progressive coalition that returns dignity into the democracy. It also includes the feminists' strike, criticises austerity as the model to combat the crisis and blaming economic powers (including the energy companies in Spain). The basic rights are understood as also those related to consumption of water or energy as the basic goods.

Figure 14 Keyword correlations in Unidos Podemos' tweets.

The co-occurrence analysis within the sentences of Podemos' tweets makes it possible to distinguish the following dimension of political persuasion: social and public, common or community´s force, political (life or vote) and feminist (fight) with the institutional and European reference.

Figure 15 Co-occurrences in Unidos Podemos' tweets.

Both of the populist parties mention social and political issues in proximity to their own party concerns (party's programme) and in correspondence to the profile of their voters. They mention a variety of the issues that form a flagship for their persuasive message while Unidos Podemos clearly tends to primarily shape its narrative against and around its main political rival, Vox. The values characteristic for both parties are present in their electoral narrative however there is a limited reference to the EU: programme, roles of the party in the EU parliament, ideas, and policies. The EU merely constitutes an institutional framework for the narrative of each party.

PSOE

When analysing the PSOE's discourse on Twitter in the EU elections it can be observed that the discourse is not as distinguished as in the case of the populist parties. First, the combinations of nouns and adjectives (the words that are following) reveals one main thematic orientation: "social justice", however without any specific proposals. The aim of the political messages is clearly persuasive as the second most prominent expression: appealing for a coherent vote. The socialist party makes a reference to the political and European project which is primarily a democratic cleaning process (understood as the elimination from the political processes or debates of right-wing populist parties such as Vox – "sanitary cordon") and pointing out the importance of the governments of the autonomous regions.

Figure 16 Nouns and adjectives combination in PSOE's discourse.

The figure also demonstrates the combination of rights and liberties associated to the party, pointing out to future advancements, for which the main protagonist referred to is the PSOE. Similarly, to Podemos, the socialists frame a discourse around the main enemy — the extreme right-wing party, portraying themselves as the only party that guarantees the future and progress of the European project.

The keywords identified by RAKE in this sense, reveals the contents of the discourse. The PSOE includes the "European project" as the mostly frequently mentioned keyword and focuses its discourse on the adversary (Vox), using words such as "sanitary cordon", "red card", "extreme right". Among the most popular political issues, the most frequently used expressions in tweets include "public housing" and "ecological transition", followed by "social cohesion" or "justice".
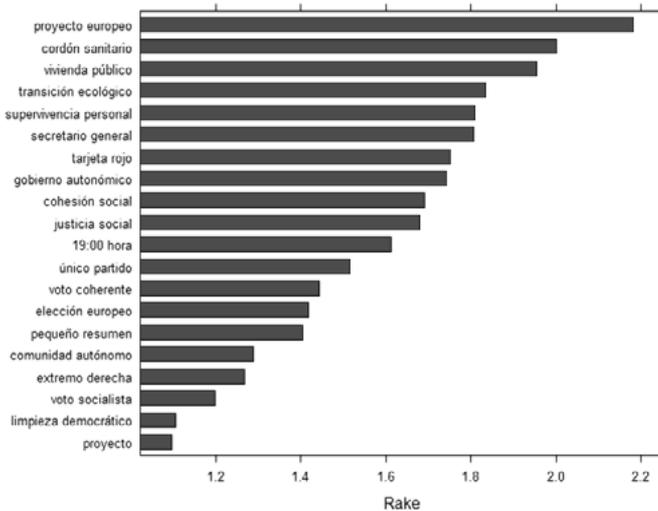

Figure 17 Most used hashtags in PSOE´s discourse.

Compared to the populist rivals, the discourse is more European-centred, but at the same time is more general, without a specific programme and is mainly built around social and progressive ideas (ecological transition), and

focused on attacks on the right-wing opponents attempting to ignore them in the political race. Keywords correlations confirm these results, adding security and unemployment as the correlating issues within the discourse and mainly institutional elements (the list, leader of the party list, general secretary, etc.). The focus is to persuade others in order to obtain a coherent vote (those supporting social progress and against the extreme right) and focused on social justice, cleaning the democratic process (from extreme right populists) and establishing social co-existence for future progress. The right wing is the main recipient of the messages, as the way to differentiate the PSOE.

PP

The PP's discourse reveals the party's programme regarding the European elections as the main topic in their tweets: job creation, concern for rural areas, and tax reform. Institutional elements of the electoral process are also advertised: campaign, rally, websites, among others. It mainly orients itself against the socialist party, as the main rival associating it with the current government. Less frequent combinations of words include: "small and medium enterprise", "new technologies", "public transport", or "concerted education", among others.
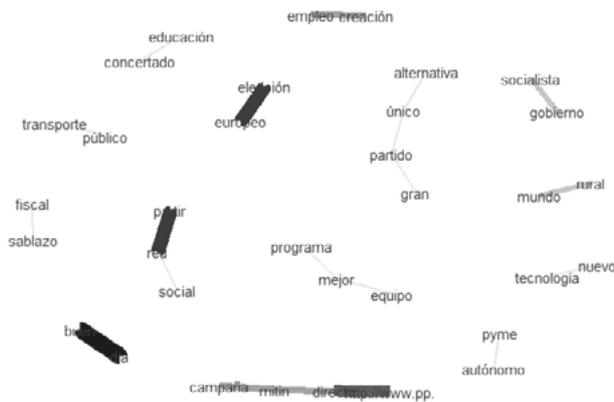
Figure 18 Words following one another in the PP's discourse on Twitter.

As a political party, the PP frames itself in a persuasive, self-affirming manner as the best programme and team, as a great party and the only alternative to its socialist counterpart in the EU elections.

RAKE analysis of the most common keywords reveals "rural areas" as the main concern shown in the messages, with the over-present reference to the "socialistic government" (the third mainly mentioned keyword). The analysis shows that it is important to consider language concerns — "good morning" is simply overused in the political message of this party and does not bring any positioning, neither help in its political branding. "European" is the third topic of the tweets, followed by a self-praising discourse. Education, women, or independentism are the issues that are far less frequently mentioned. Persuasive expressions such as "change" or "project" are those least used.
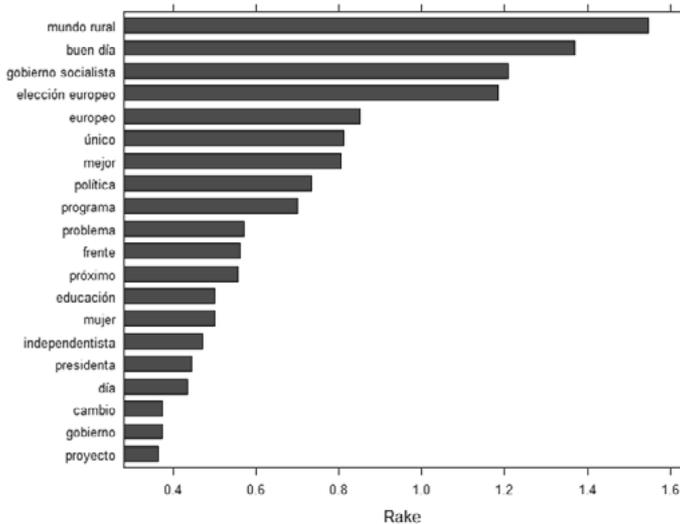

Figure 19 Most popular keywords used by PP.

The correlations of keywords bring new dimension to the discourse: above all the focus on small companies and self-employed people, together with nationalism and populism as the main threat. Therefore, on the one hand

there is a clear orientation towards economic issues and on the other a concern with populist and nationalist parties as the main opponents in the political race. Additionally, the correlations show that job creation is framed in terms of opportunities and equality.


Figure 20 Co-occurrences analysis of sentences in PP's tweets.

Analysis of co-occurrences within the sentence confirms the self-praising attitude with persuasive focus on the political process (presenting candidates, talking about debates and votes). Although the discourse refers to the European elections, it mainly emphasises the democratic and liberty-based values of the party and its economic programme (taxes). References to opponents ("left-wing", "votes for leftists") and nationalisms and populism form a significant part of the PP's narrative. Populism and nationalism are framed as equal threats and both as the principal adversaries of the PP.

Ciudadanos

Ciudadanos is the liberal party situated in the centre of the Spanish political spectrum. Ciudadanos' main discursive strategy was a strong self-appraisal as the party that is the best and only choice to fix Spanish politics in the

EU and it presented itself as such in the EU elections. First, there is a strong focus on the autonomous regions and terrorism. The party directs itself to Spanish citizens, mentions "justice" and "dignity", talks about public money and large families as the points of the programme that distinguishes it from other parties. The messages mention the political situation and the media narrative. Nationalism and populism are other two topics that can be identified among the words that follow each other.

Figure 21 Nouns and adjectives combination in the tweets of Ciudadanos.

Analysis of keywords confirms the tendency to present itself as "orange rescue" (orange is the party's colour and "orange rescue" refers to lifejackets) choosing a persuasive tone that primarily promises good management. Social and economic issues as mentioned above, plus lower taxes, are less mentioned among the published messages. Correlations of keywords reveal a use of ironic and simple, easy to understand language regarding the PSOE's management and government, that Ciudadanos considers to be deficient ("hands in the pockets", "legal trickery").

Figure 22 Co-occurrences within the sentences in the tweets of Ciudadanos.

Both analyses — co-occurrences and correlations of keywords — observe a growing dominance and presence of populism and nationalism in the party's political discourse. The overall discourse reflects a libertarian orientation towards the economy, some differentiation in terms of political positioning regarding the programme (large families, terrorism). The main focus is placed on persuasion, presenting the party as the best political choice that is the only option to rescue the country, with no clear reference to its tasks within the EU parliament.

Conclusions

The study of the 2,052 tweets published throughout the EU Parliamentary elections in Spain confirms the utility of applying text mining analysis based on computational techniques to studies of political discourse. The detailed linguistic and semantic analysis helps to reveal the focus of the political message, as well as its sentiment and tone. The frequency research is helpful to determine the dominating dimensions and to re-frame certain narratives or modify the message in order to include a more meaningful content. It is not only useful to determine the main framework of the political discourse in general during the elections or between the political rivals in order to evaluate the quality of the political debate and public opinion.

It also helps determine the scope and sentiment of the discourse and the degree of openness and freedom of political debate: its limits, main topics, malfunctions, and so forth. The computational analysis of textual data is not error-free and can be subject to further improvements. It must take into the consideration the specific features of each language and it is clear that multiple studies must be conducted in order to eliminate a random linguistics structure and to train algorithms. Furthermore, the algorithms must be adjusted to the specificity of each language. It is also of great importance to run multi-angle analysis, using different techniques and approaches: frequency, keywords, co-occurrences, words following, semantic combinations, among others. Only in this manner can the analysis fully reflect the content and dimensions of the discourse. Using a big data approach, through analysis via computational linguistic techniques of a large set of Twitter textual data, it is possible to formulate valid recommendations concerning agenda setting and framing. It also can be used to identify better linguistic expressions, more effective hashtags or keywords, and richer and better structured content.

As such, on the one hand text mining methods have made it possible to determine that the dominant issues in the Spanish electoral discourse were social questions, anti-right-wing rhetoric, and focus on Spanish political problems. On the other hand, the comparative study demonstrated that while the big parties (PSOE, PP) kept their traditional rhetoric, the messages of populists indicated several specific issues to be resolved. However, the EU approach was almost absent in the campaign.

As was observed in the case of the PP's discourse, it can be used to improve the choice of the content and to frame messages with more effective expressions. Similarly, as noticed in the case of Ciudadanos, it may be effective in modifying the persuasive messages, for example diminishing self-appraisal, in favour of more project-oriented affirmations. In the case of the PP or Ciudadanos, it can be observed that formal expressions such as "good morning" dominate the content and has no further semantic use and therefore this can be eliminated in order to benefit more meaningful expressions or keywords.

In general, the analysis revealed polarisation and fragmentation of the discourse: there is a prevailing discourse against political rivals among almost all the political parties, especially in the case of the PSOE. The PSOE's narrative is especially significant since it calls for substantial changes in the political debates, includes isolating right-wing adversaries. Additionally,

the discourse focuses primarily on Spanish politics and presents almost no reference to the EU as the political framework: projects, initiatives or ideas. While both populist parties are able to mark their political positioning in regard to their programme and values (Unidos Podemos and Vox), others apply a persuasive tone with strong auto-appraisal and branding as the only political choice, with no clear political positioning. Vox is the party with the most specific content, presenting messages that reflect the party's programme and with no references to other political rivals. Unidos Podemos, on the other hand, is rather focused on the discourse against its rival on the right wing of the political spectrum. The PSOE is the only party that presents a clear dimension of its political narrative: social justice and rights, which runs throughout the messages.

In general, the discourse of the parties during the EU elections is richer in content and more focused on the issues that dominate current politics. The general discourse of Spanish politics in the EU is rather flat, with no significant differentiation or focus. Nevertheless, the focus is placed on Spanish politics without any clear reference to EU politics, the role of the political parties in EU parliament, legislative projects, and so forth. This is probably due to the fact that the campaign for EU elections was conducted simultaneously with local authority elections in Spain.

The overall sentiment of the campaign's discourse in Spain is on the one hand very social and on the other oriented towards the future and progress, focused on polarisation and conflict with political opponents. Nationalism and populism are two notable tendencies that frequently appear throughout the tweets, constituting clear concerns among the political parties running in the EU elections. The main focus is placed on domestic politics rather than the EU itself. The dominant tone of the discourse is persuasive, encouraging voting and presenting the parties in the best possible way. There is only one slogan-hashtag of the campaign that has been captured in the frequency analysis within the analysed tweets: that of Podemos — the voter as the promoter of change.

The persuasive, general messages prevail over specific programmes during the Spanish political campaign for the EU elections. The populist parties managed to create a different positioning and the PSOE shaped its branding as the party oriented to the European project (although without clarifying it) and social progress.

Analysis of the parts of the speech that were mainly used shows that the discourse was focused on concepts and not actions, since the application

of nouns prevails. Among the most frequently used nouns the narrative points towards the future, the party as the political actors, directing towards Spanish citizens and framing the discourse around the country, freedom and change with a strong persuasive component ("change", "vote", "party"). There is no mention of the EU. "Right-wing" is one of the most frequently mentioned nouns and it is therefore possible to conclude that it dominated the general political discourse during the EU parliamentary elections. The self-appraisal attitude dominates since "better" or "the best" is the most frequently used adjective. "Social" as the adjective appears as the second most frequently use, indicating a clear social dimension of the discourse, oriented towards the near future and emphasising community ("together"). "Doing", "having" or "must be doing something" are the verbs most frequently used in the tweets, although this does not indicate any concrete or more specific action. Another verb frequently used indicates the persuasive tone — "to vote". Keywords analysis confirms the concern about right-wing popularity and participation in the elections (the most frequently used keyword is "sanitary cordon", in reference to the right-wing populist party, Vox). Among the keywords the most visible narrative is that of the socialist and left-wing Unid@s Unidos Podemos, dominated by the ecological transition, civil society or institutional issues, apart from the Catalan case ("preventive prisoner"). The latter seems to have little relation to EU issues. In general, the keywords used by the PSOE and Unidos Podemos were the most prevailing issues in the overall political discourse in Spain during the EU elections (social issues important for the left-wing in general, discourse against the right-wing party, and so forth). PMI analysis moreover reveals the personalisation of the campaign, based on the candidates and using the names of the politicians or businessmen to emphasise the narrative. Whereas the populist parties based their narrative on communicating their political proposals, the PSOE and the PP, which until now have been the biggest parties in Spain, decided to choose political positioning or, to be more precise, a persuasive strategy based on basic rhetorical devices. On the one hand there is a focus on the polemical dispute with the right-wing rival (the PSOE against Vox above all). On the other hand, the PP and Ciudadanos choose to use a persuasive style to present their parties in the best possible manner, using a self-appraisal rhetoric. Simple noun phrases reveal the popularity of the PSOE's hashtags as the only ones apart from those of Unidos Podemos that were reflected in the Spanish electoral discourse in general, and were very persuasive in their tone: vote, you can change. Some elements of self-branding and institutional branding are also present: the PSOE's "always moving forward" hashtag and the name of Partido Popular

party as the third most frequently mentioned noun combination. In this case, the reference to the EU is more clearly present. In general, the PSOE and then Unidos Podemos were stronger in the general discourse and more effective in pursuing the social agenda of their political communication, including persuasive hashtags and keywords. However, the general discourse lacks an EU orientation and is mainly focused on Spanish issues and polemical questions involving the right-wing parties. Centrist and right-wing parties did not manage an extensive presence in the general discourse and chose a rhetorical persuasive tone (PP, Ciudadanos). Populist parties (Unidos Podemos) were those which communicated their political programme and Vox was precisely, and against the assumptions, the party whose communication was primarily focused on the party's values and vision without positioning itself in the narrative against its political rivals. By contrast, Ciudadanos decided to point out the failures of their opponents and use everyday language and expressions to reach wider audiences. Nevertheless, the discourse is fragmented in terms of issues and topics in general and in the case of each party. Persuasive tone and polarisation tendency does not help enrich the contents of the general political discourse or discuss the possible solutions among the candidates and political parties. As such, the Twitter discourse of the political parties in Spain can be the subject of further improvements in terms of content and tone.

References

Adams, A., & McCorkindale, T. (2013). Dialogue and transparency: A content analysis of how the 2012 presidential candidates used twitter. *Public Relations Review, 39*(4), 357–359. https://doi.org/10.1016/j.pubrev.2013.07.016

Allagui, I., & Breslow, H. (2016). Social media for public relations: Lessons from four effective cases. *Public Relations Review, 42*(1), 20–30. https://doi.org/10.1016/j.pubrev.2015.12.001

Baines, P. R. (1999). Voter segmentation and positioning. In B. I. Newman (Ed.), *Hanbook on political marketing* (pp. 403–422). Sage Publications.

BNOSAC. (n.d.). *Basic Analytical Use Cases II: UDPipe for Topic Modelling*. NLP with R and UDPipe. https://bnosac.github.io/udpipe/docs/doc6.html

Browning N., & Sweetser K. D. (2020). How media diet, partisan frames, candidate traits, and political organization-public relationship communication drive party reputation. *Public Relations Review, 46*(2), 1–17. https://doi.org/10.1016/j.pubrev.2020.101884

Bulovsky, A. (2019). Authoritarian communication on social media: The relationship between democracy and leaders' digital communicative practices. *International Communication Gazette*, *81*(1), 20–45. https://doi.org/10.1177/1748048518767798

Campbell, S. W., & Kwak, N. (2011). Political involvement in "mobilized" society: The interactive relationships among mobile communication, network characteristics, and political participation. *Journal of Communication, 61*(6), 1005–1024. https://doi.org/10.1111/j.1460-2466.2011.01601.x

Casero-Ripollés, A., Sintes-Olivella, M., & Franch, P. (2017). The populist political communication style in action: Podemos's issues and functions on Twitter during the 2016 Spanish general election. *American Behavioral Scientist, 61*(9), 986–1001. https://doi.org/10.1177/0002764217707624

Church, K., & Hanks, P. (1990). Word association norms, virtual information, and lexicograph. In D. Samaras (Ed.), *Computational linguistics* (pp. 22–29). MIT press.

Cwalina, W., & Falkowski, A. (2015). Political branding: Political candidates positioning based on inter-object associative affinity index. *Journal of Political Marketing, 14*(1–2), 152–174. https://doi.org/10.1080/15377857.2014.990842

Deligiaouri, A. (2018). Discursive construction of truth, ideology and the emergence of post--truth narratives in contemporary political communication. *International Journal of Media & Cultural Politics, 14*(3), 301–315. https://doi.org/10.1386/macp.14.3.301_1

Dong, C., & Ji, Y. (2018). Connecting young adults to democracy via government social network sites. *Public Relations Review, 44*(5), 762–775. https://doi.org/10.1016/j.pubrev.2018.05.004

Druckman, J. N. (2014). Pathologies of studying public opinion, political communication, and democratic responsiveness. *Political Communication, 31*(3), 467–492. https://doi.org/10.1080/10584609.2013.852643

Echeverría, M. (2017). Political personalisation and journalistic infotainment. A frame analysis. *Cuadernos.Info,* (41), 71–86. https://doi.org/10.7764/cdi.41.1099

Enli, G., & Skogerbø, E. (2013). Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. *Information, Communication & Society, 16*(5), 757–774. https://doi.org/10.1080/1369118X.2013.782330

Erzikova, E., & Bowen, S. A. (2019). Missed opportunities to make PR great again: How public relations agencies responded to the Trump Presidency. *Public Relations Review, 45*(5), 1–8. https://doi.org/10.1016/j.pubrev.2019.05.014

European Parliament. (2019, July 2). *Constitutive session - 23/10/2019.* 2019 European election results. https://www.europarl.europa.eu/election-results-2019/en

Feldman, R., & Sanger, J. (2006). *The text mining handbook.* Cambridge University Press.

Fontecilla Camps, M. E. (1988). La "semantización" del discurso político. *Revista Comunicación y Medios,* (7–8), 47–52. https://dialnet.unirioja.es/servlet/articulo?codigo=5242724

Frame, A., & Brachotte, G. (2015). Le tweet stratégique: Use of Twitter as a PR tool by French politicians. *Public Relations Review. 41*(2), 278–287. https://doi.org/10.1016/j.pubrev.2014.11.005

Froehlich, R., & Rüdiger, B. (2006). Framing political public relations: Measuring success of political communication strategies in Germany. *Public Relations Review, 32*(1), 18–25. https://doi.org/10.1016/j.pubrev.2005.10.003

Gálvez-Rodríguez, M., Sáez-Martín, A., García-Tabuyo, M., & Caba-Pérez, C. (2018). Exploring dialogic strategies in social media for fostering citizens' interactions with Latin American local governments. *Public Relations Review, 44*(2), 265–276. https://doi.org/10.1016/j.pubrev.2018.03.003

Guo, L., Mays, K., & Wang, J. (2019). Whose story wins on Twitter? Visualising the South China Sea dispute. *Journalism Studies*, *20*(4), 563–584. https://doi.org/10.1080/1461670X.2017.1399813

Hong, H. (2013). Government websites and social media's influence on government-public relationships. *Public Relations Review, 39*(4), 346–356. https://doi.org/10.1016/j.pubrev.2013.07.007

Housholder E., & LaMarre H. L. (2015). Political social media engagement: Comparing campaign goals with voter behaviour. *Public Relations Review, 41*(1), 138–140. https://doi.org/10.1016/j.pubrev.2014.10.007

Jansen, A. S., Eugster, B., Maier, M., & Adam, S. (2019). Who drives the agenda: Media or parties? A seven-country comparison in the run-up to the 2014 European Parliament elections. *International Journal of Press-Politics, 24*(1), 7–26. https://doi.org/10.1177/1940161218805143

Kelm, O., Dohle, M., & Bernhard, U. (2019). Politicians' self-reported social media activities and perceptions: Results from four surveys among German parliamentarians. *Social Media and Society, 5*(2), 1–12. https://doi.org/10.1177/2056305119837679

Kiousis, S., & Shields, A. (2008). Intercandidate agenda-setting in presidential elections: Issue and attribute agendas in the 2004 campaign. *Public Relations Review, 34*(4), 325–330. https://doi.org/10.1016/j.pubrev.2008.07.004

LaMarre, H. L., & Suzuki-Lambrecht, Y. (2013). Tweeting democracy? Examining Twitter as an online public relations strategy for congressional campaigns'. *Public Relations Review, 39*(4), 360–368. https://doi.org/10.1016/j.pubrev.2013.07.009

Lawless, B., & Chen, Y. (2019). Developing a method of critical thematic analysis for qualitative communication inquiry. *Howard Journal of Communications, 30*(1), 92–106. https://doi.org/10.1080/10646175.2018.1439423

Lee, J., & Lim, Y. (2016). Gendered campaign tweets: The cases of Hillary Clinton and Donald Trump. *Public Relations Review, 42*(5), 849–855. https://doi.org/10.1016/j.pubrev.2016.07.004

Lee, J., & Xu, W. (2018). The more attacks, the more retweets: Trump's and Clinton's agenda setting on Twitter. *Public Relations Review, 44*(2), 201–213. https://doi.org/10.1016/j.pubrev.2017.10.002

Macnamara, J. (2012). Public communication practices in the web 2.0-3.0 mediascape: The case for PRevolution. *PRism, 7*(3), 1–13. https://www.prismjournal.org/uploads/1/2/5/6/125661607/v7-no3-a3.pdf

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market--structure surveillance through text mining. *Marketing Science, 31*(3), 521–543. https://doi.org/10.1287/mksc.1120.0713

Nicholls, T., & Bright, J. (2019). Understanding news story chains using information retrieval and network clustering techniques. *Communication Methods and Measures*, *13*(1), 43–59. https://doi.org/10.1080/19312458.2018.1536972

Painter, D. L. (2015). Online political public relations and trust: Source and interactivity effects in the 2012 U.S. presidential campaign. *Public Relations Review, 41*(5), 801–808. https://doi.org/10.1016/j.pubrev.2015.06.012

Pedersen, T. L. (2021, February 23). *An implementation of grammar of graphics for graphs and networks*. https://cran.r-project.org/web/packages/ggraph/ggraph.pdf

Plowman, K., Wakefield, R., & Winchel, B. (2015). Digital publics: Tracking and reaching them. *Public Relations Review*, *41*(2), 272–277. https://doi.org/10.1016/J.PUBREV.2014.12.007

Pressgrove, G., & Kim, C. (2018). Stewardship, credibility and political communications: A content analysis of the 2016 election. *Public Relations Review, 44*(2), 247–255. https://doi.org/10.1016/j.pubrev.2018.01.003

Reynolds, C. (2019). Building theory from media ideology: Coding for power in journalistic discourse. *Journal of Communication Inquiry, 43*(1), 47–69. https://doi.org/10.1177/0196859918774797

Roper, J., & Hurst, B. (2019). Public relations, futures planning and political talk for addressing wicked problems. *Public Relations Review, 45*(5), 1–8. https://doi.org/10.1016/j.pubrev.2019.101828

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In M. W. Berry & J. Kogan (Eds.), *Text mining: Applications and theory* (pp. 3–20). John Wiley & Sons.

Rune, K., & Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with twitter data. *The International Journal of Press/Politics, 21*(3), 338–357. https://doi.org/10.1177/1940161216645335

Sahly, A., Shao, C.., & Kwon, K. H. (2019). Social media for political campaigns: An examination of Trump's and Clinton's frame building and its effect on audience engagement. *Social Media and Society*, *5*(2), 1–13. https://doi.org/10.1177/2056305119855141

Sisco, T. S., Lucas, J. C., & Galdieri, C. J. (2017*). Political communication and strategy: Consequences of the 2014 midterm elections*. The University of Akron Press.

Speakman, B. (2015). Interactivity and political communication: New media tools and their impact on public political communication. *Journal of Media Critiques, 1*(1), 131–144.

Stetka, V., Surowiec, P., & Mazák, J. (2019). Facebook as an instrument of election campaigning and voters' engagement: Comparing Czechia and Poland. *European Journal of Communication, 34*(2), 121–141. https://doi.org/10.1177/0267323118810884

Tameryan, T. Y., Zheltukhina, M. R., Slyshkin, G., Abakumova, O., Volskaya, N., & Nikolaeva, A. (2018). Metaphor in political media discourse: Mental political leader portrait. *Online Journal of Communication and Media Technologies, 8*(4), 377–384. https://doi.org/10.12973/ojcmt/3958

Trimithiotis, D. (2018). Understanding political discourses about Europe: A multilevel contextual approach to discourse. *Discourse & Society, 29*(2), 160–179. https://doi.org/10.1177/0957926517734425

Vesnic-Alujevic, L. (2012). Political participation and web 2.0 in Europe: A case study of Facebook. *Public Relations Review, 38*(3), 466–470. https://doi.org/10.1016/j.pubrev.2012.01.010

Wang, J. (2016). A new political and communication agenda for political discourse analysis: Critical reflections on critical discourse analysis and political discourse analysis. *International Journal of Communication, 10,* 2766–2784. https://ijoc.org/index.php/ijoc/article/view/3847/1679

Watkins B. A. (2017). Experimenting with dialogue on Twitter: An examination of the influence of the dialogic principles on engagement, interaction, and attitude. *Public Relations Review, 43*(1), 163–171. https://doi.org/10.1016/j.pubrev.2016.07.002

Wijffels, J. (2021, June 1). *Tokenization, parts of speech tagging, lemmatization and dependency parsing with the 'UDPipe' 'NLP' toolkit.* https://cran.r-project.org/web/packages/udpipe/udpipe.pdf

Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication, 59*(1), 40–56. https://doi.org/10.1111/j.1460-2466.2008.01403.x

Xiong, Y., Cho, M., & Boatwright, B. (2019). Hashtag activism and message frames among social movement organisations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Public Relations Review. 45*(1), 10–23. https://doi.org/10.1016/j.pubrev.2018.10.014

Xu, S., & Xiong, Y. (2020). Setting socially mediated engagement parameters: A topic modeling and text analytic approach to examining polarized discourses on Gillette's campaign. *Public Relations Review, 46*(5), 1–12. https://doi.org/10.1016/j.pubrev.2020.101959