

Article

Not peer-reviewed version

Psicolinguistic Modeling of Spanish Political Parties from Social Media

[Arturo Montejo-Ráez](#)*, [Manuel Carlos Díaz-Galiano](#), [Luis Navarro-Ardoy](#),
Manuel Tomás González-Fernández

Posted Date: 25 October 2023

doi: 10.20944/preprints202310.1671.v1

Keywords: Social Computing; Topic Modeling; Psicolinguistics; Natural Language; NLP



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Psicolinguistic Modeling of Spanish Political Parties from Social Media

Arturo Montejo-Ráez ^{1,*}, Manuel Carlos Díaz-Galiano ², Luis Navarro-Ardoy ³
and Manuel Tomás González-Fernández ⁴

¹ CEATIC, Universidad de Jaén, 23071 Jaén, Spain

² CEATIC, Universidad de Jaén, 23071 Jaén, Spain; mcdiaz@ujaen.es

³ Sociology Department, Universidad Pablo de Olavide, 41013 Sevilla, Spain; lnavard@upo.es

⁴ Sociology Department, Universidad Pablo de Olavide, 41013 Sevilla, Spain; mtgonfer@upo.es

* Correspondence: amontejo@ujaen.es

Abstract: This study applies several natural language processing techniques for characterizing social groups by analyzing published text in social media. The resulting computed values are proposed as features for modeling social groups and measure distances or similarities among them, along further data that could help understanding group behavior according to language usage. In this study, we have analyzed the messages published on Twitter regarding patriotism by the main political parties in Spain. The results show that lexical resources and topic modelling algorithms are useful to model different groups and serve as comparison tools to better understand the topics these groups are talking about.

Keywords: social computing; topic modeling; psicolinguistics; natural language; NLP

1. Introduction

Today, it is difficult not to agree that words are the raw material of politics. Without them there is no discourse, no project [1]. When the topics that enter the political debate concern issues related to the Spanish nation, terminology is very important. In general, it is so both because of the relationship that right-wing and left-wing ideologies have with nationalism and, in particular, because of the particular developments that the idea of the Spanish nation has had in our country [2].

Therefore, the words that political parties choose to designate a reality are not arbitrary. It is not the same to talk about Spain using terms referring to its political-administrative structure (territorial organisation) as it is to include more affective terms or visions (national sentiment, patriotism...) [3,4]. For researchers, the huge amount of data circulating on the Internet has become a very rich field of analysis. Big Data is a new product of the knowledge society made possible by digital technologies, social networks and cloud computing [5].

In this article we attempt to characterise the political parties at the national level (*PP*, *PSOE*, *Podemos*, *C's*, *UPyD* and *IU*) based on the analysis of texts about Spain published on the popular social network Twitter (currently known as X). Using various techniques, we propose the calculation of characteristics that can establish measures of similarity or distance between these parties along with other elements that allow us to know their behaviour at the level of language management.

2. Background

The use of social networks for opinion analysis has become one of the most active areas of research in recent years [6], consolidating as additional services offered by companies that provide real-time reports and dashboards on various information flows, such as Onclusive¹ or Metricool².

¹ Available online: <https://onclusive.com/> (accessed on 23 October 2023)

² Available online: <https://metricool.com/> (accessed on 23 October 2023)

In Spain, projects such as AORESCU [7] have been carried out, focused on the analysis of user opinions in social networks for the Spanish language, which involves the generation of resources and the study of algorithms. The work of [8] involves a detailed review of the studies carried out on the use of social networks (especially Twitter/X) for political analysis, with the prediction of electoral results as a main objective. From this review we extract that we face a complex problem, which deserves to be studied in depth, and with several elements involved: group bias, gender dominance, ideological relationship between users, etc.

Recommending products based on user profiles is something that involves the generation, first, of such profiles [9], which are usually built from the interactions of each user and identifying certain elements (such as the tags used). The study of user profiles in social networks is of great interest to the industry, as it allows information to be routed to recipients with a higher acceptance rate. In fact, there are already some patents in this regard, such as Google's patent for generating a list of events of interest to users of a social network [10].

In our case, the focus is on characterising *groups*. A recent and outstanding work in this regard is that of [11]. In this work, they study data collected from the Twitter social network to analyse which factors lead to political change based on the social traits detected. With 87% accuracy, they conclude that, after highlighting the cultural battles between ideologies in the case of the proposal of a law for same-sex marriage, it is possible to build a statistical model that predicts the potential for political change. In this sense, they empirically observed that positions that were more emotionally charged and smaller in number rarely led to such change. This is an example of the possibility of modelling ideologies from big data.

3. Data Preprocessing

For this study, publications on the popular social network Twitter were collected for 18 days, not always consecutive, between the months of October and November 2015. The days monitored respond to the context of the "unprecedented" general elections of 20D. Unprecedented because all the polls predicted the end of the two-party system and the entry into the political arena of emerging groups such as *Ciudadanos* and *Podemos*. Data were collected before the 20D elections (collected in the pre-election period) and after the 20D elections (collected in the post-election period). Specifically, for the month of October, the 20th, 21st, 28th, 29th, 30th and 31st were monitored. For November, from the 1st to the 10th, on the 12th and from the 19th to the 24th. The specific days as well as the hours of capture have responded to the availability of the network and the storage infrastructure. The data was extracted through the Twitter API³ using Logstash⁴ as extractor over Elasticsearch⁵ as storage.

The Spanish bag of keywords used to retrieve tweets is as follows:

"organización territorial del estado, sentimiento de identificación con españa, estado de las autonomías, identidad nacional, estado autonómico, patria, reforma constitucional, patriotismo, federalismo, sentirse español, nacionalismo español, patriotas, nación, país, constitución, estructura, historia, pluralidad de naciones, lengua, autogobierno, competencias, identidad española, soberanía, etnia, independencia, raza, separación, cultura, comunidades autónomas, herencia cultural, nacionalidad, costumbres, descentralización, tradiciones, constitución de 1978, comunidad imaginada, subjetivo, identidades nacionales, españolista, independentista, autonomista, genes, singularidad, soberanismo, catalanismo, pueblo, sentimiento nacionalista, bandera, himno".

These words translated into English are:

"territorial organisation of the state, feeling of identification with spain, state of autonomies, national identity, autonomous state, homeland, constitutional reform, patriotism, federalism, feeling Spanish, Spanish nationalism, patriots, nation, country, constitution, structure, history,

³ Available online: <https://developer.twitter.com/en/docs/twitter-api> (accessed on 23 October 2023)

⁴ Available online: <https://www.elastic.co/logstash> (accessed on 23 October 2023)

⁵ Available online: <https://www.elastic.co/> (accessed on 23 October 2023)

plurality of nations, language, self-government, competences, Spanish identity, sovereignty, ethnicity, independence, race, separation, culture, autonomous communities, cultural heritage, nationality, customs, decentralisation, traditions, 1978 constitution, imagined community, subjective, national identities, Spanish nationalism, independence, autonomism, genes, singularity, sovereignty, Catalan nationalism, people, nationalist sentiment, flag, anthem".

The tweets were filtered and processed as follows:

1. Only textual information of the tweet was retained (no additional metadata is considered).
2. Only tweets with a minimum of 6 terms were considered.
3. Retweets were discarded
4. Tweets with links were discarded
5. Text was normalised to lowercase, without accents and compressing letter repetitions to two occurrences.

The statistics of the tweets obtained after filtering and processing can be seen in Table 1.

Table 1. Corpus statistics

Days	Tweets	Data Size	Vocablary Size
28	5,530,927 tweets	554 MB	159,587 terms

4. Tools for Psycholinguistic Characterization

Our aim is to **characterize social groups through language**. To this end, we apply two methods that allow us to extract certain measures of language use, at the lexical, syntactic and semantic levels, for the generation of vectors that serve as models of the different political parties: Analysis of psycholinguistic features with the Linguistic Inquiry and Word Count (LIWC) system and the representation of the groups by Topic Modeling. Some of these models will not allow a certain interpretation of language use, others only provide us with an abstract representation at a statistical level for the generation of these vectors.

4.1. Psycholinguistic Features of LIWC

As Zhang and Counts [11] did, we are going to use the word-based psychometric indicators defined in the LIWC 2007 dictionary, in its Spanish version, and developed by LIWC. It is actually a very simple program: it reads a text and calculates the percentage of words related to different emotions, ways of thinking, social traits and even some grammatical aspects. The base dictionary has been developed by researchers with interests in social, clinical, health and cognitive psychology, so the language categories are designed to capture the psychological and social states of individuals [12]. The base dictionary was provided to us by James Pennebaker himself and we have built our own application on it for use in text analysis. The Spanish dictionary is the result of a translation and revision of the 2007 LIWC in English and was carried out by two Mexican experts and a Spanish expert. This version contains 7,515 words and word roots, with the same number of categories and subcategories as the English LIWC [13].

The measures established to reflect *ideology* fall under the following categories:

- **Openness:** Clifford and Jerit [14] in their study indicates that of the Big 5 Personality Features (Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness) from *openness* and *conscientiousness* correlate with ideology. They use the work of Yarkoni [15] to match LIWC with openness through correlations with their categories. Correlation values with *openness* greater than 0.2 (which is the threshold that gives Zhang and Counts the best results), positive or negative are: *total pronouns* (-0.21), *articles* (0.2), *time* (-0.22), *motion* (-0.22) and *grooming* (-0.2), although the latter has been removed in the current versions of LIWC. We do not know if these correlations are

applicable to Spanish, so we assume this possible bias and propose to carry out future research similar to that of [Yarkoni](#) for Spanish language.

- **Emotions and feelings:** they use LIWC to measure *positive* and *negative* emotion according to the content of the texts as well as the prevalence of emotions such as *anger* and *anxiety*, together with the prevalence of *swear words* (*Maldec* in Spanish LIWC).
- **Certain:** *Certain* and *tentative* measures are added.
- **Bipartisanship:** In addition to these measures that we can select from the literature, we have also decided to include the categories *ppron* and *they* as indicators of *bipartisanship* as we understand that the continuous references to other parties reflect a use of language oriented towards argumentation based on an opponent.

Morality, as [Zhang and Counts](#) understand it, would correspond to the categories *harm*, *fairness*, *purity*, *ingroup* and *authority* which, they add [16]. Also include the prevalence of the term *religion*. Unfortunately, this dictionary is not available for Spanish, so we have not considered this category as a feature of political ideology.

Table 2 summarises these correspondences between the different versions of LIWC by language:

Table 2. Matching between LIWC and Spanish LIWC.

Feature	LIWC	Spanish LIWC
Openness	total pronouns	TotPron, -0.21
	articles	Articulo, 0.2
	time	Tiempo, -0.22
	motion	Movim, -0.22
Emotions and feelings	affec	Afect
	posemo	EmoPos
	negemo	EmoNeg
	anger	Enfado
	sad	Triste
	anx	Ansiedad
Certain	swear	Maldec
	certain	Certeza
Bipartisanship	tentative	Tentat
	ppron	ProPer
	they	Ellos(Ellas)

However, despite this selection, we will study how valid it is to take all LIWC categories to differentiate psycholinguistic features and which ones provide more information, generating alternative vectors that allow us to observe whether the characterisations by selecting some LIWC categories are different or not from when we take all of them.

5. Topic Modeling

Topic modeling is a task that aims to extract the "topics" of a given set of documents [17]. Several algorithms that try to solve this, but all of them are based on calculations on the probability distributions of words, sequences of words and sets of documents. One of the most widely used algorithms is Latent Dirichlet Allocation (LDA) [18], which considers that each document is a mixture of a small number of topics, and that the inclusion of each word is related to one of the topics covered in the document. This is very useful for analysing groups of people based on the texts they generate, as is the case in our study, because not only can we represent each group as a vector of topics (with associated weights), but we can also visualise the most significant words in each topic, which can give us an idea of the topics that are of interest to each group.

Twitter topic tracking has provided interesting works, such as that of [19], which generates a social graph from a seed topic. This question is related to an area of Natural Language Processing

(NLP) called Topic Detection and Tracking (TDT). Our interest is in extracting topics from the message history of different social groups (political parties, in this case), and not so much in tracking the topics discussed.

6. Experiments

6.1. Linguistic Inquiry and Word Count (LIWC)

The selected features are those listed in Table 3, providing vectors of 15 elements, which are given the following values for each match according to the calculation on the LIWC patterns:

Table 3. Measured Spanish LIWC values for the 15 selected categories. They form, by columns, the vectors of ideological characteristics of each party. The correspondence with the original LIWC categories can be seen in Table 2.

	C's	IU	Podemos	PP	PSOE	UPyD
Openness						
TotPron	0,164228	0,171207	0,179671	0,197391	0,191635	0,193254
Articulo	0,197265	0,194188	0,191063	0,206423	0,196706	0,201728
Tiempo	0,010781	0,014057	0,025757	0,029475	0,021154	0,048754
Movim	0,010859	0,022507	0,024291	0,030905	0,022793	0,016398
Emotions and feelings						
Afect	0,050269	0,045250	0,045106	0,031084	0,040244	0,030375
EmoPos	0,032568	0,023533	0,028228	0,021474	0,031701	0,021791
EmoNeg	0,017975	0,024086	0,015245	0,009431	0,011898	0,007484
Enfado	0,012570	0,011135	0,008879	0,004839	0,005457	0,001816
Triste	0,003460	0,006712	0,002220	0,001553	0,002507	0,000666
Ansiedad	0,001163	0,002764	0,002639	0,001663	0,001427	0,001101
Maldec	0,000479	0,002685	0,001759	0,000880	0,000829	0,000165
Certain						
Certeza	0,008230	0,017373	0,019936	0,026616	0,011107	0,011115
Tentat	0,031336	0,021322	0,036521	0,023852	0,035327	0,012766
Bipartisanship						
PronPer	0,104321	0,109295	0,114629	0,124060	0,133593	0,115776
Ellos	0,031424	0,018874	0,024291	0,034837	0,021057	0,019534

Using the column of values as the vector representing the political party, a matrix of distances between parties can be obtained. Figure 1 graphically shows the similarity (1 minus the distance) between parties. Through a principal components analysis (PCA) on each category using the parties as variables (to study the variation of language based on the parties), we found that with only two principal components we could explain 98% of the variance. Taking the first two components it is possible to show in a two-dimensional space the distance between parties, represented by the cosine distance, as shown in Figure 2.

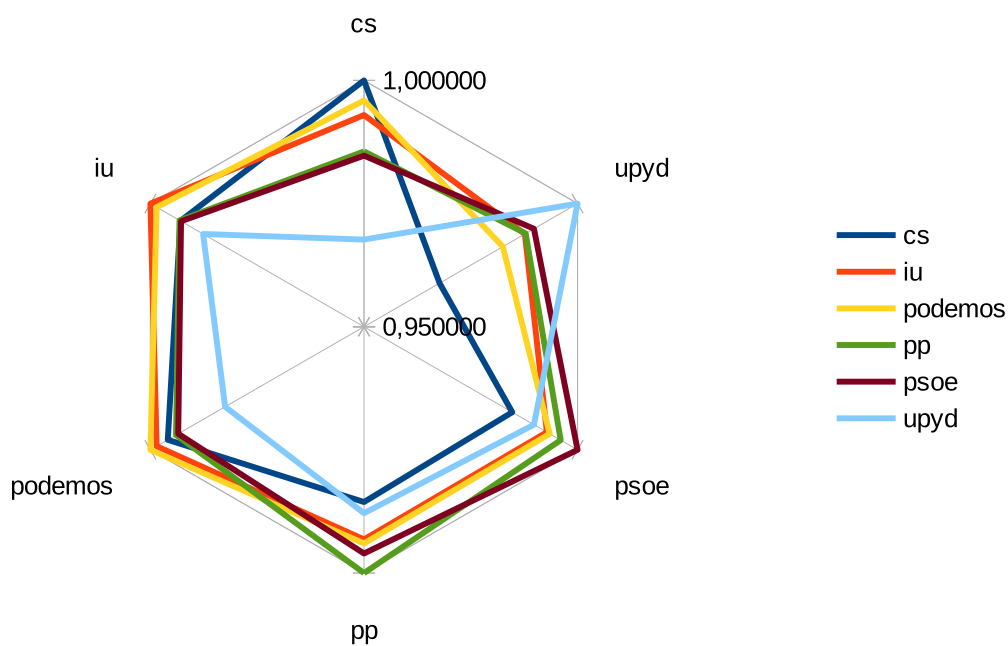


Figure 1. Similarity between political parties on the basis of the 15 ideological features identified with LIWC.

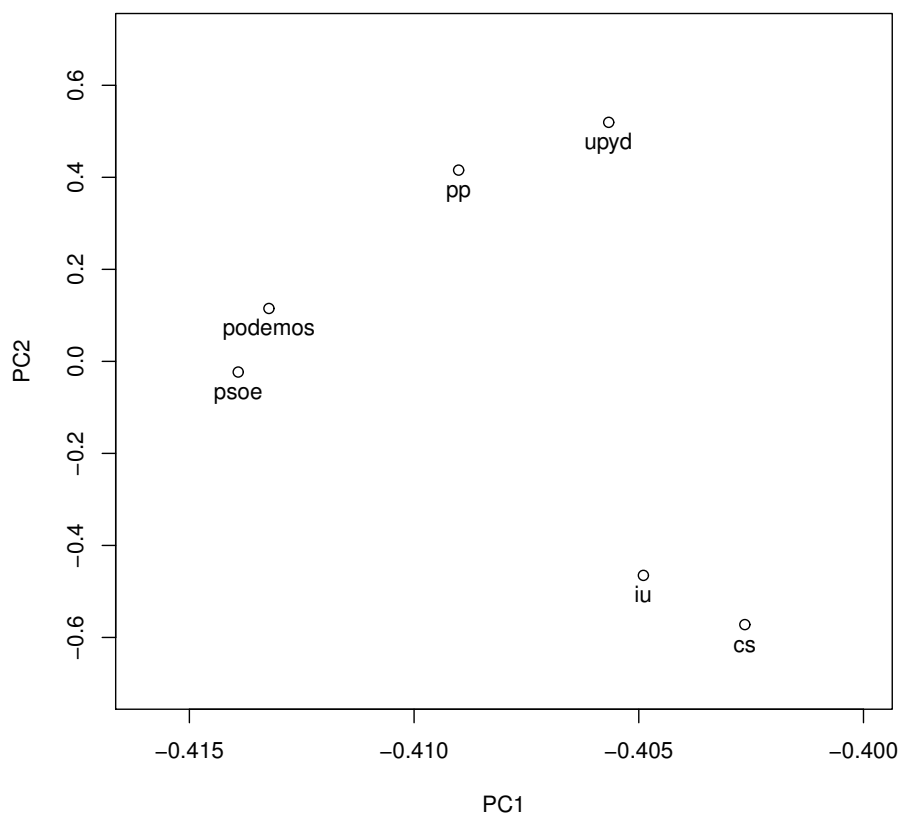


Figure 2. Two-dimensional representation of the projected match space on its two main components.

6.2. Topic Modeling

The LDA method [18] is a generative model that assumes that each document is a mixture of a small number of categories and the occurrence of each word in a document is due to one of the categories to which the document belongs. This algorithm is similar to *Probabilistic Latent Semantic Analysis* [20] and is widely used for topic-based text characterization.

For the calculation of LDA, the Topic Modeling module of the MALLET⁶ (Machine Learning for Language Toolkit) software has been used. Table 4 shows the 10 most relevant topics extracted by LDA, with the most representative terms associated with them. We have added a fourth column indicating the party or parties for which the weight of this issue is relevant in their discourse. As mentioned above, the topics can also serve as dimensions as a way of group vectorisation. Figure 3 shows graphically the weights of each topic for each political party.

⁶ Available online: <http://mallet.cs.umass.edu/topics.php> (accessed on 23 October 2023)

Table 4. Words associated with each topic calculated with LDA and the parties closest to that topic (in bold the topic that would most characterise the party). The third column shows a translation into English of the Spanish terms.

Topic	Spanish Terms	Translated Terms	Parties
0	pueblo ahora cataluna soberania pues gente nunca nacion mayor voto politico dinero catalunya cosas elecciones corrupcion	people now catalonia sovereignty so people never nation biggest vote political money catalonia things elections corruption	<i>Podemos</i>
1	historia independencia espana mejor independentista espanol hablar mismo referendum parte gran alguien lengua catalanes hoy hombre proceso	history independence spain best independentist spanish speak same referendum part great someone language catalans today man process	<i>C's and IU</i>
2	historia bandera cataluna patria pais frente partidos usted golpe viva catalana leyes parece pagar verguenza mayoria mismo patriotas declaracion	history flag catalonia homeland country front parties you coup viva catalana laws seems pay shame majority same patriots declaration	<i>PP</i>

Table 4. Cont.

Topic	Spanish Terms	Translated Terms	Parties
3	presidente espana ninguno instrumentos juridicos objetivos alcance utilizara lograran garantizo nacion juntos democracia espanoles europea preservaremos avanzada fractura logrado	president spain none instruments legal objectives scope scope will_use will_achieve guaranteed nation together democracy spanish_people european will_preserve advanced fracture achieved	<i>PP</i>
4	cataluna independencia rajoy acabo manana deje prensa rueda financiar personalmente anunciar constitucion autogobierno autonomia golpistas emana suspendido espanola votar	catalonia independence rajoy finish tomorrow let press conference finance personally announce constitution self-government autonomy coup_plotters emanate suspended spanish voting	<i>UPyD</i>
5	pais constitucion cultura partido politica siempre espanola queremos anos personas psoe menos gobierno quieren espanoles quiero ley votar	country constitution culture party politics always spanish we_want years people psoe less government they_want spanish_people I_want law vote	<i>PSOE and Podemos</i>

Table 4. Cont.

Topic	Spanish Terms	Translated Terms	Parties
6	hijos lider gusta historia segun acuerdo pregunten tdt espana reforma religion laico bienestar vida constitucional psoe	children leader like history according_to agreement ask tdt spain reform religion secular welfare life constitutional psoe	<i>PSOE</i>
7	pais contigo gente independencia cultura gracias cambio favor proteger dia menor abrazos grupos democracia disfrute participe ministro	country with_you people independence culture thanks change please protect day lower hugs groups democracy enjoyment participate minister	<i>Podemos</i>
8	izquierdas programa eliminar impuestos defendiendo laico educacion electricas concertada progresivos nacionalizar rajoy generales elecciones gobierno patriotas pueblo gane empenado cerrazon	left-wing programme eliminating taxes defending secular education electricity concerted progressive nationalise rajoy general elections government patriots people win insist bloody-mindedness	<i>IU</i>

Table 4. Cont.

Topic	Spanish Terms	Translated Terms	Parties
9	anos	years	C's
	democracia	democracy	
	negros	blacks,	
	dictadura	dictatorship	
	franco	franco	
	representa	represents	
	cuidemos	care value	
	valoremos	died anthem	
	murio himno	country	
	pais empresas	companies	
	armas saudi	weapons	
	trilero	saudi	
	vendera	swindler	
	arabia	will_sell	
	esplendor	arabia	
	maximo	splendour	
	historia	maximum	
		history	

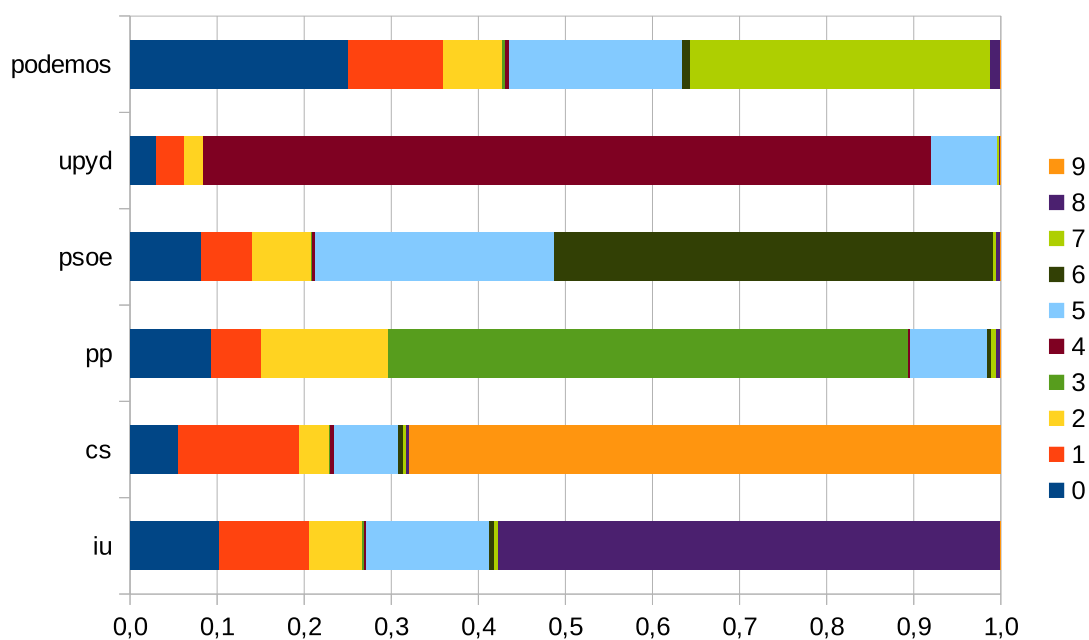


Figure 3. Weight of each topic in each party

7. Analysis

As can be seen, it is possible to use the proposed techniques to study the discourse of the parties, with interesting visualisations that serve as a tool for analysis. The proximity between *Podemos* and *PSOE* in terms of patriotic discourse is remarkable, as is the proximity between *IU* and *Ciudadanos*. We also observe that *UPyD* presents the greatest differentiating feature over the rest of the parties. Thanks to topic modelling, we observe that this party focuses its discourse on the Catalan issue.

8. Conclusions and Future Work

This paper presents two proposals for the characterisation of social groups based on posts made on the popular social network Twitter/X. Both allow the generation of identifying features, using psycholinguistic categories or terms grouped into topics. These two techniques make it possible to understand the model generated, as both the categories and the themes have a specific semantics.

However, from the point of view of a vector space as a form of representation, it would be possible to explore options such as word vectors based on Milokov's method of bags of words and n-grams [21]. This method has been applied to handle the semantics of words in problems such as term clustering or analogy search.

In addition, we could approach the analysis of the parties at a more global level, and not focused on the use of Spanish national sentiment in the discourse. This is why we are generating a broader corpus that will allow us to generate more generic profiles. Likewise, it is necessary to validate the information generated with respect to other studies carried out in sociology, in order to verify the concordance between the conclusions drawn with traditional instruments and those obtained through the analysis of macro-data.

Author Contributions: Conceptualization, A.M. and L.N.; methodology, A.M.; data curation, A.M. and M.C.D.; software, M.C.D. and A.M.; validation, L.N. and M.T.G.; formal analysis, A.M. and L.N.; investigation, A.M., M.C.D., L.N. and M.T.G.; resources, M.C.D. and A.M.; writing—original draft preparation, A.M., M.C.D., L.N. and M.T.G.; writing—review and editing, A.M., M.C.D., L.N. and M.T.G.; visualization, A.M.; supervision, A.M.; project administration, A.M.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by Spanish Ministry of Economy and Competitiveness (project TIN2015-65136-C2-1-R).

Institutional Review Board Statement: Not applicable

Data Availability Statement: The anonymized data presented in this study are available on request from the corresponding author. The complete data are not publicly available due to privacy, because it was obtained from Twitter/X and are available at <https://api.twitter.com> with the permission of X Corp.

Abbreviations

The following abbreviations are used in this manuscript:

PP	<i>Partido Popular</i>
PSOE	<i>Partido Socialista Obrero Español</i>
UPyD	<i>Unión Progreso y Democracia</i>
IU	<i>Izquierda Unida</i>
C's	<i>Ciudadanos</i>
API	Application Programming Interface
LIWC	Linguistic Inquiry and Word Count
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
TDT	Topic Detection and Tracking
PCA	Principal Component Analysis

References

1. Gutiérrez-Rubí, V. El silencio en política, 2016. [Online; 29 Marzo de 2016].
2. Ferri-Fuentevilla, E.; Ruiz-Jiménez, A.M. Entre patria y estado: Formas de nombrar España. Un recorrido por los discursos programáticos de PSOE y AP-PP entre 1977 y 2011. *Empiria. Revista de metodología de ciencias sociales* **2015**, *32*, 63–84.
3. Easton, D. *A systems analysis of political life*; Wiley New York, 1967.
4. Anderson, B.; Suárez, E.L. *Comunidades imaginadas: Reflexiones sobre el origen y la difusión del nacionalismo*; Fondo de Cultura Económica México, 1993.
5. González de la Fé, T. Sociología y Big Data. *Encrucijadas. Revista crítica de ciencias sociales* **2014**, *8*, 51–53.

6. Martínez-Cámara, E.; Martín-Valdivia, M.T.; Urena-López, L.A.; Montejo-Ráez, A.R. Sentiment analysis in Twitter. *Natural Language Engineering* **2014**, *20*, 1–28.
7. Troyano Jiménez, J.A.; Ureña López, L.A.; Maña López, M.J.; Cruz Mata, F.; Enríquez de Salamanca Ros, F. AORESCU: Análisis de opinión en redes sociales y contenidos generados por usuarios. *Procesamiento del Lenguaje Natural* **2015**, *55*, 153–156.
8. y David Vilares, M.A.A. A review on political analysis and social media. *Procesamiento del Lenguaje Natural* **2016**, *56*, 13–24.
9. Guy, I.; Zwerdling, N.; Ronen, I.; Carmel, D.; Uziel, E. Social media recommendation based on people and tags. In Proceedings of the Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2010, pp. 194–201.
10. Bosworth, A.G.; Cox, C.; Sanghvi, R.; Ramakrishnan, T.S.; D'angelo, A. Generating a feed of stories personalized for members of a social network, 2010. US Patent 7,827,208.
11. Zhang, A.X.; Counts, S. Modeling Ideology and Predicting Policy Change with Social Media: Case of Same-Sex Marriage. In Proceedings of the CHI, 2015, pp. 2603–2612.
12. Pennebaker, J.; Chung, C.; Ireland, M.; Gonzales, A.; Booth, R. The development and psychological properties of LIWC2007, 2014.
13. Ramírez-Esparza, N.; Pennebaker, J.W.; García, F.A.; Suriá Martínez, R.; et al. La psicología del uso de las palabras: Un programa de computadora que analiza textos en español. *Revista Mexicana de Psicología* **2007**, *24*.
14. Clifford, S.; Jerit, J. How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics* **2013**, *75*, 659–671.
15. Yarkoni, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality* **2010**, *44*, 363–373.
16. Graham, J.; Haidt, J.; Nosek, B.A. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology* **2009**, *96*, 1029.
17. Wallach, H.M. Topic modeling: Beyond bag-of-words. In Proceedings of the Proceedings of the 23rd international conference on Machine learning. ACM, 2006, pp. 977–984.
18. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *the Journal of machine Learning research* **2003**, *3*, 993–1022.
19. Cotelo, J.M.; Cruz, F.L.; Troyano, J.A. Dynamic topic-related tweet retrieval. *Journal of the Association for Information Science and Technology* **2014**, *65*, 513–523.
20. Hofmann, T. Probabilistic latent semantic indexing. In Proceedings of the Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1999, pp. 50–57.
21. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *CoRR* **2013**, *abs/1301.3781*.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.