A Pertinence Score for Political Discourse Analysis: The Case of 2018 Colombian Elections

JUAN CARLOS GARCIA, FIEC, Escuela Superior Politecnica del Litoral, Guayaquil, Ecuador and Facultad de Ingeniería Industrial, Universidad de Guayaquil, Guayaquil, Ecuador
DAMIÁN FERNÁNDEZ PEDEMONTE, Austral University, Buenos Aires, Argentina and National Scientific and Technical Research Council, Buenos Aires, Argentina
CARMEN VACA, Facultad de Ingeniería en Electricidad y Computación (FIEC), Escuela Superior Politecnica del Litoral, Guayaquil, Ecuador
MARIA BAQUERIZO-NEIRA, Universidad Casa Grande, Guayaquil, Ecuador

This study proposes a quantitative method to assess the pertinence of political language on national issues, addressing the complexity of analyzing political discourse and its relevance to citizens' concerns. Using word embeddings and linguistic models trained on Wikipedia, a "pertinence score" was developed to measure the relevance of political discourse in contexts such as the economy and health. The method was applied to the 2018 Colombian presidential election, revealing significant differences in thematic pertinence between candidates. Survey validation confirmed the correlation between automatic and human scores, highlighting the model's ability to discriminate ideological positions through lexical analysis.

CCS Concepts: • **Computing methodologies → Discourse, dialogue and pragmatics;**

Additional Key Words and Phrases: Political discourse, neural language model, elections, Colombia

ACM Reference Format:
Juan Carlos Garcia, Damián Fernández Pedemonte, Carmen Vaca, and Maria Baquerizo-Neira. 2024. A Pertinence Score for Political Discourse Analysis: The Case of 2018 Colombian Elections. *Digit. Gov. Res. Pract.* 5, 3, Article 34 (October 2024), 15 pages. https://doi.org/10.1145/3689213

1 Introduction

Political discourse analysis is crucial for decision-making in democratic societies, as it clarifies citizens' understanding of complex issues and helps them form reasoned judgments about potential solutions [27]. Analyzing political discourse can be challenging due to the complexity of topics like education, health, and the economy,

Juan Carlos García also with Universidad de Guayaquil, Facultad de Ingeniería Industrial.
Damián Fernández Pedemonte also with National Scientific and Technical Research Council, CONICET.
Authors' Contact Information: Juan Carlos Garcia, FIEC, Escuela Superior Politecnica del Litoral, Guayaquil, Guayas, Ecuador and Facultad de Ingeniería Industrial, Universidad de Guayaquil, Guayaquil, Guayas, Ecuador; e-mail: jgarcia@espol.edu.ec; Damián Fernández Pedemonte, Austral University, Buenos Aires, Buenos Aires, Argentina and National Scientific and Technical Research Council, Buenos Aires, Buenos Aires, Argentina; e-mail: dfernandez@austral.edu.ar; Carmen Vaca, Facultad de Ingeniería en Electricidad y Computación (FIEC), Escuela Superior Politecnica del Litoral, Guayaquil, Guayas, Ecuador; e-mail: cvaca@fiec.espol.edu.ec; Maria Baquerizo-Neira, Universidad Casa Grande, Guayaquil, Guayas, Ecuador; e-mail: gbaquerizo@casagrande.edu.ec.

ACM 2639-0175/2024/10-ART34
https://doi.org/10.1145/3689213

and the vast amount of information that citizens need to process. While previous research has primarily focused on classifying political discourse based on ideological leanings [13, 52], or examining its linguistic structures [1], there is a need to understand how discourse relates to broader contextual and structural frameworks [60, 62]. To address this gap, we introduce a novel methodology utilizing language models based on word embeddings [39, 48]. Our approach quantifies the relevance of political discourse concerning issues of national importance, thereby facilitating a more objective and reliable evaluation of how well political discussions align with citizens' concerns.

Speeches by political leaders during campaigns are often filled with promises. While conventional democratic theories expect the assessment of these promises to be balanced, the reality often shows an asymmetrical dynamic. For instance, the 'cost of governing' suggests that voters are more inclined to penalize leaders for unmet promises than to reward them for fulfilled ones [41]. Moreover, voters frequently forget intricate details and are sometimes unaware of fundamental facts about their political systems and representatives [58].

This highlights the challenge democracies face in providing accessible and high quality information resources to help citizens evaluate the relevance of policies proposed in a given political discourse. Quantitative measures of the significance of candidates' proposals could aid voters in making more informed decisions. Current research points to a general disinterest among the Latin American electorate in candidates' proposals, with a greater focus on their public images [5, 6]. Some scholars argue that political unawareness is a prevalent trait among democratic citizens [14, 19, 35, 55], yet a well-informed electorate is crucial for a functioning democracy [9, 32]. This paper aims to contribute to that goal by quantifying the relevance of a political discourse regarding the expectations of the population in a given country.

Contrary to supporting the agenda-setting hypothesis, which is less applicable in the age of social media [2], our study suggests that in electoral contexts, candidates might still influence the agenda when compelled to respond to unforeseen events [37]. However, our primary interest lies in the discourse crafted by candidates concerning public interest issues.

"Pertinent" in our study is defined as highly-related to an issue of decisive or significant relevance that aids in understanding the subject under discussion. We have developed a pertinence score to quantify how significantly political discourse addresses issues such as economic crises, violence, or health. Moreover, this score is encapsulated by the most relevant terms that signify its essence.

To evaluate the pertinence of language in political discourse, we propose a method that scrutinizes the lexicon and extracts key terms using advanced techniques such as Word Embeddings [39, 48]. Our constructed model offers citizens an objective means of comparing the proposals of different candidates, thereby facilitating a more informed and analytical assessment of political rhetoric. This approach not only enhances transparency in political communication but also empowers voters by providing them with a data-driven tool to evaluate the relevance and substance of political discourse.

Our contributions are fourfold:

- First, we construct a set of categories listing the main concerns selected by Latin American citizens in a well-known survey (Latinobarómetro). These categories reflect prevalent issues in Latin America such as high rates of corruption, violence, crime, and increasing poverty and unemployment [12, 53], which are analyzed in the discourse of political leaders.
- Second, we introduce and apply a 'Pertinence Score' to evaluate political discourse on specific issues like the economy, corruption, and education, based on the semantic meaning of the words.
- Third, using word embedding, we create language models applicable to diverse contexts, including political discourse and general contexts like Wikipedia, allowing for a nuanced measurement of discourse pertinence.
- Finally, we apply this relevance score to the 2018 Colombian presidential elections, which garnered significant attention following Petro's victory in 2022.

2 Latinobarómetro: Public Opinion in Latin America

Since the mid-1970s, the rise of democracy has notably increased freedom, subsequently catalyzing the development of public opinion polls. These polls, often driven by researchers and political organizations, aim to study voter behavior during election campaigns. By the mid-1990s, systematic public opinion polling had expanded internationally, becoming a pivotal tool in the social sciences for monitoring social and political changes globally.

The Latinobarómetro is a comparative study that tracks the evolution and transformation of societies, focusing on public policy-oriented democracies [51]. Regional barometers, including the Asian Barometer, Arab Barometer, Afrobarometer, Eurasia Barometer, Eurobarometer, and the Latin Barometer, provide insights across different geographical contexts.

The Latinobarómetro project, managed by the nonprofit Latinobarómetro Corporation, is a public good accessible through a data bank. Annually, it surveys approximately 20,000 citizens from 18 Latin American countries on various aspects such as democracy, institutions, economy, social media, science, trust, and international relations.

3 Language and Political Communication

Numerous studies have constructed profiles of political leaders by analyzing the natural language in their speeches [46, 47, 49]. Content analysis has been used to create political dictionaries and predict outcomes of parliamentary elections, such as in Venezuela using tweets from political candidates and their audiences [15].

Research has also explored the use of argumentative strategies and new discursive dynamics on X (formerly known as Twitter) to boost electoral success [43, 45, 61]. A study on the Spanish party Podemos examined X messages to understand the populist communication style of leader Pablo Iglesias, revealing his use of anti-elitist language [20].

Some researchers propose Bayesian human-interpretative language models to identify political topics, noting that candidates use social networks like Facebook and X for various purposes. These studies have found no correlation between the issues candidates prioritize and those that concern citizens [56]. Another study differentiated between figurative language and literal meanings, using computational models to detect sarcasm on X [8]. However, contributions to understanding the pertinence of language in political discourse, particularly in Latin America, are limited.

This study aims to delve into aspects of political communication not typically addressed, moving beyond marketing strategies to focus on the rhetorical techniques employed by candidates. This approach helps assess how closely the topics candidates discuss align with public demands and societal issues.

The application of linguistic analysis to large samples of text from social media is still in its early stages. It primarily focuses on word repertoires or topics, under the premise that recurring lexical items can reveal political priorities and predict public impact. Semantic analysis techniques, which include collecting semantic markers that group words into fields and measure the distance between them, aid in transitioning from lexical to textual analysis, enhancing coherence and identifying ideologically charged language. Our approach aims to quantify the linguistic relevance of each candidate within different thematic categories, supporting the hypothesis that more ideologically charged words tend to have indirect denotations and intense connotations, such as labeling a violent event as a “massacre.”

4 Computational Methods for Text Analysis

The use of automated text analysis techniques on social media content represents a dynamic research area. Recent developments in topic discovery across political data have utilized unsupervised topic modeling techniques like **Latent Dirichlet Allocation (LDA)** [11] and **Non-negative Matrix Factorization (NMF)** [16], which effectively reveal prevalent topics and associated vocabularies. These techniques are particularly valuable in agenda-setting contexts, where it is crucial to recognize the influence of political campaigns and candidates on shaping public perceptions of pressing issues. Although this research does not directly focus on agenda-setting theory proposed by McCombs and Shaw in 1972 [38], the impact of political discourse on public concerns is evident.

Natural language processing (NLP) techniques have shown the capability to discern patterns in discourse that align with agenda-setting strategies [23, 26]. Nonetheless, our methodology emphasizes the direct evaluation of thematic relevance from the citizens' perspective.

Studies on polarization and preference identification through online posts have provided insights into political orientations [7, 10]. Traditional text classification methods, which employ machine learning and deep learning, have been widely used for sentiment analysis [57] and stance detection [54], although they struggle to capture the nuanced semantic relationships of words. Recently, neural network-based word embeddings [24, 39, 48] have shown promise in distinguishing semantic similarities, with text classifiers like CNNs [31] and **RNNs (Recurrent Neural Networks)** [17] achieving superior performance over older methods.

Our analysis seeks to measure the lexical pertinence of political discourse themes, assessing the significance each candidate assigns to various topics. By analyzing the pertinence across different subjects for each candidate's discourse, we can infer the ideological nuances reflected in their language choices. This leads to several hypotheses:

(1) The pertinence of language in candidates' political discourse on social media is an identifiable and measurable indicator of the importance attached to each contextual category and the ideological differences between candidates.
(2) Semantic relationships in language use can determine the pertinence of political discourse.
(3) It is feasible to construct a pertinence score from a language model that focuses on semantic word relationships within specific contexts.

5 Methods and Experiments

This section describes the approach used to define the pertinence score of the political discourse. First, the categories are determined based on the needs of Latin American public opinion. Second, the data is collected from the politicians' accounts on X. Then, the language model is built from the collected data. Finally, the relevance score is calculated from the most significant words in each category. These steps are detailed in the following sections.

5.1 Categories Definition

We structure our analysis of political discourse on X around a set of categories pertinent to the Latin American context. To establish these categories, we utilize data from the Latinobarómetro, which focuses on 'personal, social, and country issues'. Specifically, we analyze responses to the question: "In your opinion, what is the most important problem in the country?" From these responses, we identify the main concerns and categorize them into relevant themes. Consequently, we have defined 12 categories:

As a result, we define 12 categories,

(1) CORRUPTION: Institutional problems.
(2) ECONOMY: Economic, financial problems, inflation, and price increases.
(3) EDUCATION: Concerns about eligibility and cost.
(4) ENERGY: Issues with public services such as water and electricity, as well as transport and fuel.
(5) ENVIRONMENT: Pollution and global warming.
(6) HEALTH: Public health problems and drug addiction.
(7) HOUSING: Issues related to housing cost and location.
(8) HUMAN RIGHTS: Human rights issues and discrimination.
(9) POVERTY: Crime, lack of security, drug trafficking, gangs, and terrorism.
(10) VIOLENCE: Crime, lack of security, drug trafficking, gangs, and terrorism.

(11) WORK: Unemployment and job insecurity.
(12) YOUTH: Youth values, future, and lack of opportunities.

5.2 Data Description

In the context of the 2018 Colombian presidential elections, our analysis focuses on the political discourse of the two finalist candidates: Ivan Duque and Gustavo Petro. The corpus for this study is derived from data accessed through the X REST API, which provides timelines of users. We collected tweets from the accounts "@IvanDuque" and "@petrogustavo" from January 1, 2018, to June 17, 2018. Our dataset includes 2,353 tweets from Ivan Duque and 3,786 from Gustavo Petro, totaling 6,139 tweets analyzed during the election campaign up to the day of the presidential election.

5.3 Data Processing

To enhance readability and facilitate further analysis, we process the textual data extensively. This involves the removal of unnecessary characters and symbols. Additionally, each tweet undergoes a tokenization process, where words are extracted for analysis. To improve efficiency and reduce noise in the data, we also remove common connector words, known as stop-words, such as "the, a, an, there, here." This cleanup process is crucial for refining the data into a usable form for subsequent analyses.

5.4 Language Model

A language model plays a crucial role in predicting word distributions, capturing both semantic knowledge and grammatical structures of the natural language used in document corpora. In our study, each tweet is treated as a separate document. Neural language models use continuous word representations to predict how 'close' or 'distant' terms are within a specific context, such as violence or corruption.

A language model has the central role of predicting word distributions, which have to capture semantic knowledge and grammatical structure in the natural language used to build documents in a corpus. In our case, each tweet is a document. Neural language models apply continuous representations of words to make predictions and learn how 'close' or 'distant' the terms are to a specific context, such as violence or corruption. The algorithm constructs a vector for each word where two 'close words' are those with very similar vectors. These word representations (embeddings) are stored in a matrix $W \in \mathbb{R}^{dx\|V\|}$ where V is the vocabulary of all unique words found in the entire set of documents.

In recent years, neural network-based language models have become increasingly significant due to their ability to effectively represent word meanings and capture the semantic properties of language [44, 50, 52]. We utilize such models to analyze the corpus detailed previously.

The primary aim of employing a neural network language model is to grasp the semantic representation of words linked to each category within the candidates' corpus. This allows us to quantitatively assess the group of words a leader uses when discussing specific topics. Interestingly, these words need not be synonyms of the category; it is sufficient that they share similar vector representations. Among the most popular algorithms for language modeling based on neural networks are Word2Vec [39] and **GloVe (Global Vectors for Word Representation)** [48].

Consider the sentence: "Liberal and progressive spirit to reach an agreement on the fundamental: Peace, Full Democracy and Productive Economy". Word2Word2Vec, which is well-suited for creating word representations in short texts like tweets, offers two models: **CBOW (Continuous Bag of Words)** and Skip-gram. CBOW predicts a target word based on context words, for example, using 'liberal' and 'spirit' to predict 'progressive'. Conversely, Skip-gram does the opposite, using a target word to predict its context, which makes it ideal for our needs, as it helps identify words that define the context of a given category.

These models are structured in layers, each serving a specific function. The Skip-gram model, for instance, includes an input layer that represents word vectors, a hidden layer that processes these vectors and predicts the probability distribution of vocabulary words, and an output layer that mirrors the input layer's dimension, predicting context words.

GloVe operates on a different principle, using global co-occurrence counts of words to capture their meanings. This model optimizes the least-squares error, producing a word vector space that retains semantic similarities. For example, the probability that a j word occurs in the context of word i is determined by:

$$P_{ij} = P(j \mid i) = \frac{X_{ij}}{X_i}.$$

where X is the co-occurrence matrix. GloVe embeddings, trained on extensive corpora, encapsulate the general meaning of words, proving invaluable in discerning semantic similarities in brief texts [28, 30].

Our ultimate objective is to develop a methodology that yields a pertinence score, quantifying the political discourse as manifested through tweets. To fulfill this, we must align the context of each category with a sufficiently general and extensive corpus, such as Wikipedia, to capture the broad meaning of words. The proximity between these two contexts allows us to accurately calculate the pertinence score, as detailed in the following section.

5.5 Pertinence Model

To initiate, we train a custom word embedding using the Word2Vec toolkit, enabling us to extract the ten most closely related words for each category and candidate. These words embody the semantic context prevalent in each leader's political discourse. However, for an accurate pertinence calculation, a broader set of standard words–such as those found in Wikipedia or Google News–is necessary. We therefore utilized a pre-trained GloVe embedding model, featuring a million words from Wikipedia, to ascertain the semantic meanings of selected words. This approach allows us to measure the distance between the lexicon used in political discourse and the relevant public issues.

The choice of Wikipedia for linguistic model training is justified by its extensive subject coverage [22], continuous updates, and accessibility [4]. As a massive and diverse corpus, Wikipedia is ideal for learning the semantic nuances of words in a general context. Its collaborative nature ensures that the content reflects the latest developments across various fields [40]. Additionally, the preprocessing of Wikipedia data for training linguistic models, as seen in question-answer system developments, highlights its utility and direct applicability in natural language processing tasks [36].

The Skip-gram model, adept at predicting the context from a set of words given a single target word, utilizes a context window defined by $-c$ and c , with the word at index t. The probability function for Skip-gram is given by:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c}\log p(w_{t+j}|w_t)$$

Cosine similarity, commonly used to assess the semantic relationship between two word embeddings, determines the semantic relatedness of words in each category. The similarity between two words i and j is projected on the similarity of the vectors v_i and v_j:

$$sim(v_i, v_j) = \cos(\theta) = \frac{v_i \cdot v_j}{\parallel v_i \parallel \times \parallel v_j \parallel}.$$

We evaluate whether the similarity of embedding words correlates with human judgments of word similarity. However, it is crucial to consider how these words, specific to a politician's rhetoric, diverge from the general definitions of public issues. Therefore, we utilize a larger and more complex corpus from Wikipedia to capture global meanings. Larger corpora typically yield better statistics and hence, more accurate word meanings. For this, we employed a pre-trained word representation based on Wikipedia, which outperforms larger corpora like Gigaword due to the constant updates capturing new knowledge. We use a publicly available pre-trained

Table 1. A Sample with Three Words Obtained Automatically by the Pertinent Model

Duque	Petro	Duque	Petro	Duque	Petro
CORRUPTION		ECONOMY		EDUCATION	
officer (0.65)	corrupt (0.82)	tax (0.80)	productive (0.64)	higher (0.76)	higher (0.76)
denounce (0.58)	injustice (0.77)	market (0.75)	farming (0.62)	preschool (0.75)	professional(0.79)
against (0.56)	impunity (0.67)	grow-up (0.72)	industry (0.57)	virtual (0.75)	quality (0.76)
ENERGY		ENVIRONMENT		HEALTH	
consumption(0.60)	solar (0.90)	flora (0.76)	energies (0.61)	prevention (0.61)	hospital (0.81)
efficient (0.50)	electric (0.87)	wildlife (0.75)	zones (0.58)	problems (0.53)	patient (0.79)
actions (0.45)	generators (0.81)	biodiversity(0.69)	existence (0.57)	education (0.52)	doctors (0.77)
HOUSING		HUMAN	RIGHTS	POVERTY	
floors (0.78)	tends (0.85)	constitution(0.70)	humans (0.82)	class (0.59)	population (0.72)
share (0.77)	marginal (0.56)	law (0.63)	constitution(0.78)	expand (0.51)	levels (0.68)
owners (0.74)	roads (0.54)	agreements(0.61)	exercise (0.73)	growing (0.51)	percentage(0.67)
VIOLENCE		WORK		YOUTH	
aggression (0.68)	exclusion (0.72)	dedication (0.65)	job (0.80)	talents (0.62)	youth (0.78)
intolerance(0.64)	ignoring (0.71)	inspiration (0.65)	posts (0.80)	parks (0.61)	education (0.73)
rejection (0.58)	slavery (0.71)	posts (0.65)	entry (0.80)	creativity (0.61)	modernity (0.71)

GloVe model computed using the “Spanish Billion Words Corpus and Embeddings” and then calculate the cosine distance to quantify the lexical distance in the candidate’s context relative to a given category.

Finally, the pertinence score is calculated as the average of the semantic distances of the most salient words in each category. For instance, in Gustavo Petro’s discourse, the word ’hospital’ is closely associated with the ’health’ category, exhibiting a semantic distance of 0.81. Conversely, for Ivan Duque, the word ’prevention’ is most pertinent in the same category, with a score of 0.61.

6 Results

Our findings confirm the hypotheses that the use of language in political discourse on social networks serves as a measurable indicator of the importance attributed to each contextual category. We operationalized each category through a list of words ranked by their relevance. Table 1 illustrates some of these relevant words, organized by category and political candidate.

The semantic relationships between words play a crucial role in determining the pertinence of language within political discourse. For example, in the context of energy, the words ‘solar (0.9)’, ‘electric (0.87)’, and ‘generators (0.81)’ are significantly more pertinent compared to ‘consumption (0.60)’, ‘efficiency (0.50)’, and ‘actions (0.45)’. These differences allow us to construct a relevance score based on the semantic relationships of words within specific contexts. Notably, candidate Gustavo Petro achieves a higher pertinence score than his competitor Ivan Duke in the energy category, while the reverse is observed in the housing category.

The subsequent section details the validation process of these results. Following that, in the discussion section, we further explore these findings in terms of ideological differentiation between the candidates.

6.1 Evaluation

To validate our methodology, we designed an experiment to obtain relevance scores based on human judgments. Individuals in the field of communication are typically adept at determining the relevance of a word to a topic. Consequently, we developed a survey that asks respondents to score words for their pertinence on a Likert scale from 1 to 5, where 1 signifies ’not pertinent’ and 5 denotes ’extremely pertinent’. This survey consists of 12 questions, each corresponding to a different category, with the names of the candidates omitted to prevent bias.

Each question requires participants to evaluate 20 words, and a total of 300 surveys were administered to undergraduate and graduate students via Google Forms. The demographic breakdown of respondents includes

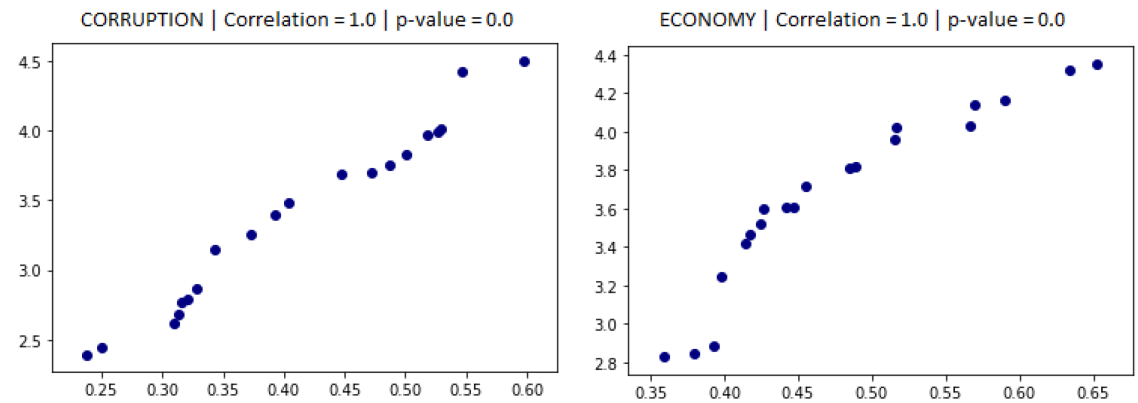

Fig. 1. Spearman's rank correlation for Corruption and Economy. The x-axis is the quantitative rank and the y-axis is the qualitative rank.

(65%) women and 81% individuals under 25 years old, majoring in fields like communication and journalism (43%), arts and multimedia design (31%), and administration and marketing (18%).

Upon completion of the surveys, we calculated the correlation between the scores given by participants and those generated by our model for each word set within the categories. The statistical measure used to evaluate the strength of this relationship is Spearman's rank correlation coefficient, defined as:

$$r_i = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}.$$

where i is the index of the category, d is the difference between the ranks for the paired observations, and n is the number of paired words, set at ten for this study. A high correlation coefficient suggests a strong agreement between the human-assigned scores and our model-generated scores, indicating that both sets of data provide similar information.

As revealed by the Figure 1, the rank correlation achieves a perfect score of 1.0 for categories such as corruption and economy. The correlation for categories like energy, environment, health, housing, human rights, poverty, and violence also reaches 1.0. However, categories such as education, work, and youth have a slightly lower correlation of 0.9.

This part of our study underscores the discursive nature of political language, which is inherently persuasive and unfolds its argumentation throughout the text. Like all genres of discourse, political language requires an understanding of context, reflected in various textual operations [62]. Political language, understood as the lexical construction by politicians, often features connotative elements [29], produces words related to shared agendas, employs polysemic or ambiguous terms, and uses euphemisms [42]. Additionally, the performative aspect of political language–the actions intended through word use–only manifests effectively within specific contexts [18].

Furthermore, metaphorical language plays a crucial role, enabling politicians to discuss one reality in terms of another, thereby defining and valuing social realities [33, 34]. Structural semantics has extensively studied this level, developing methodologies to analyze the interconnected universe of signs, independent of their contextual appearance [25].

Our research quantifies word usage within context-specific categories including corruption, economy, education, and others, demonstrating the pertinence of each candidate's discourse within these categories. As illustrated in Figure 2, our analysis reveals distinct differences in the relevance attributed to various topics by Duque and Petro, with notable disparities in categories like energy, poverty, and rights. This analysis helps contrast

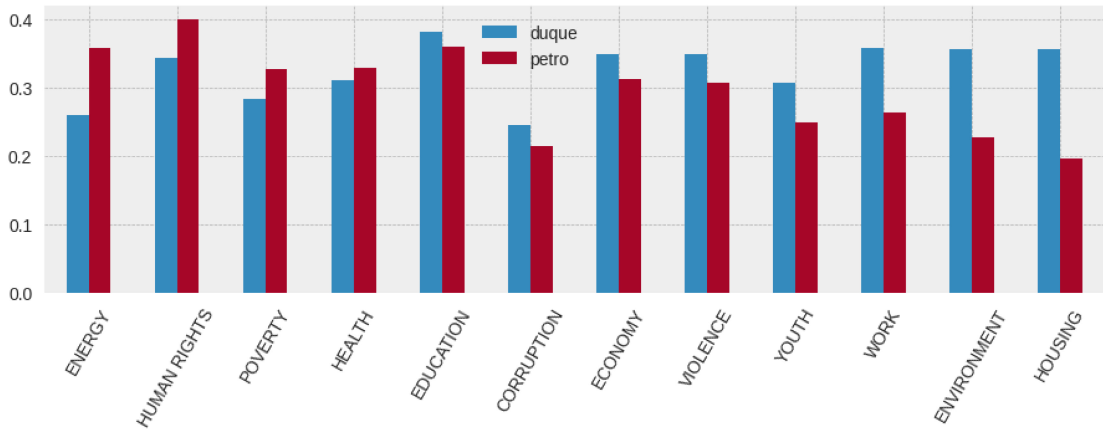

Fig. 2. Ivan Duque and Gustavo Petro's political speech pertinence score calculated for each category.

Fig. 3. On the left, Iván Duque's proposal on X regarding Education. On the right, Gustavo Petro's proposal on X regarding Energy.

the linguistic strategies employed by each candidate with their respective political propositions, as seen in their campaign manifestos in Figure 3.

In the following sections, we will delve deeper into these comparisons, analyzing how the choices made by political leaders align with their campaign proposals and the broader societal contexts they aim to influence.

6.2 Categories with Different Scores among the Candidates

We first analyze the categories of Housing and Violence, where Duque's scores exceed those of Petro. Subsequently, we explore categories where the opposite trend is observed.

6.2.1 Housing. In examining the lexical choices where the two political actors diverge, Duque demonstrates greater pertinence than Petro in the Housing category. Not only do the degrees of pertinence differ, but so do the connotations of the ten most pertinent words. For Duque, the top words include 'houses', 'flats', 'homes', 'rural', 'credit', 'own', 'dignified', 'financing', 'owner', and 'term'. These terms suggest a narrative focused on facilitating access to housing ('credit', 'financing,' 'term'), promoting ownership ('own,' 'owner'), and ensuring dignity ('dignified'). The inclusion of 'rural' broadens the scope of this rights-based approach.

By contrast, Petro's top words —'pay', 'be', 'roads', 'results', 'gave', 'banks', 'businessman', 'certain', 'equal', and 'discount'— are more ambiguous and could apply to various contexts, such as Tourism. This ambiguity suggests that, isolated or grouped, these terms do not clearly align with a semantic axis focused on housing.

Reviewing the original tweets from Duque's account provides concrete examples of programmatic language aimed at housing improvements and access:

> "We will improve 600,000 houses in the next 4 years (bathrooms, kitchens, flats, utilities) and create a new housing program with no down payment where what you pay for rent is the rent of your own house #CorazonGrande #DuqueConLaComunidad #AguaBlanca #Cali."[1]

> "We'll create a new housing program with no down payment, where what you pay for rent is the down payment on a homeowner's #DuqueEsElQueEs #DuqueConLosTrabajadores."[2]

> "In #Acevedo #Huila we manifest our rural housing improvement program. 180,000 houses improved (floors, roofs, kitchens and bathrooms) to improve the quality of life in the countryside."[3]

6.2.2 Violence. Similarly, in the Violence category, Duque and Petro not only differ in the degree of pertinence but also in the connotations of their chosen terms. Duque's terms —'aggression', 'threat', 'woman', 'intolerance', 'abandonment', 'expression', 'result', 'type', 'rejection', 'absence'— directly refer to aspects of violence. The inclusion of 'woman' adds a dimension of gender violence.

Petro's selections —'injustice', 'inequality', 'exclusion', 'slavery', 'laws', 'reconciliation', 'love', 'segregation', 'third', 'ignorance'— employ a more metaphorical use of language to depict violence. Terms like 'injustice' and 'exclusion' represent abstract forms of violence, while 'reconciliation' and 'love' suggest antidotes to violence.

Tweets from Petro's account reinforce his engagement with the theme of gender violence, often framing it as a significant issue within his political narrative:

> "Why @AlvaroUribeVel evades talking about the @ClaMoralesM violation. Violence against women is a matter of power. Democracy implies the equality and dignity of Women or it is not Democracy."[4]

> "The aggression against our candidate to the Senate on the DECENTES party, @AidaAvellaE, surviving victim of the genocide against the UP, is simply an act of cowardice. They can't stand a worthy woman, they can't stand the difference, they can't stand ideas, they can't stand love."[5]

This analysis not only highlights the lexical choices made by each candidate but also underscores the ideological underpinnings of their discourse. Petro's emphasis on combating gender violence is seen as part of a broader ideological narrative that aligns with progressive values on gender equality.

In this context, it becomes evident that Petro's political discourse, more so than Duque's, actively promotes the fight against gender violence. This issue is not only a prominent feature of his political platform but also serves as a distinctive ideological statement within his broader political narrative [3, 62]. Given that discussions on gender are often more ideologically charged than other topics [59], Petro's emphasis on combating gender

[1] https://twitter.com/IvanDuque/status/964679528428826625
[2] https://twitter.com/IvanDuque/status/970780542600966144
[3] https://twitter.com/IvanDuque/status/959905631523229699
[4] https://twitter.com/petrogustavo/status/956353345635127296
[5] https://twitter.com/petrogustavo/status/962298979022229504

violence aligns with a ‘motivating-programmatic’ approach within his campaign. This commitment is part of his governmental proposals, through which he advocates for the cause. Moreover, his focus on gender issues reflects a left-wing ideological stance that prioritizes equality and rights, specifically gender parity in this instance [3]. This ideological orientation underscores a broader commitment to addressing issues of inequality and promoting social justice within his political agenda.

6.2.3 Energy. In the Energy category, Petro demonstrates greater pertinence than Duque. Petro’s ten most relevant terms include: ‘electric’, ‘solar’, ‘gas’, ‘generated’, ‘clean’, ‘generators’, ‘installation’, ‘use’, ’ ‘solar’, and ‘cost’. Duque’s top terms are: ‘consumption’, ‘efficient’, ‘cost’, ‘high’, ‘transform’, ‘national’, ‘past’, ‘council,’ ‘contact’, and ‘rate’. A significant portion of Petro’s most pertinent words are qualifying adjectives (‘clean’, ‘electric’, ‘solar’) or terms related to relevant technology (‘solar’, ‘generators’), indicating a technical understanding of diversified and non-conventional energy sources. Petro’s vocabulary belongs to an environmentalist lexicon. Conversely, Duque’s vocabulary is less technical, focusing on traditional state energy management issues such as ‘consumption’, ‘high cost’, and ‘rate’.

Example of Duque’s tweet:

> “#IvánDuqueEnRCNRadio| Micro, small and medium enterprises are suffocated by the tax burden, so I propose to reduce this burden and also have differential tariffs. @rcnradio”.[6]

Example of Petro’s tweet:

> “The energy of Colombia Mrs. @ladossa does not come from petroleum, but fundamentally from water; its hydroelectric plants generate 66% of the total energy, I propose to progressively replace the electric energy generated in thermoses based on gas and coal by clean solar energy”.[7]

In the Violence category, it is crucial to correct the “false adscription of relevance” [21] through detailed analysis. For example:

> “#Florence | The affection of the citizens of Caquetá fills me with energy and commits me to work for a better country #FutureFromAll”[8]

6.2.4 Poverty. In the Poverty category, Petro’s terminology is more relevant than Duque’s. Petro’s top terms include: ‘reduction’, ‘live’, ‘generalized’, ‘population’, ‘levels’, ‘gap’ ‘percentages’, ‘zones’, ‘major’, and ‘decrease’. Duque’s terms are: ‘vulnerable’, ‘objective’, ‘childhood’, ‘home’, ‘nutrition’, ‘class’, ‘achieved’, ‘expansion’, ‘above’, and ‘medium’. Petro’s choice of words better defines the contours of the poverty issue and emphasizes his commitment to addressing it, using terms like ‘generalized’, ‘levels’ ‘zones’ and ‘gap’. His use of action-oriented verbs such as ‘reduction’, ‘live’, ‘reduce’ and ‘decrease’ adds a proactive tone to his discourse. In contrast, Duque’s vocabulary is more bureaucratic and localized, with terms like ‘vulnerable’ and ‘childhood,’ and he frequently references the family as a victim of poverty.

Example of Petro’s tweet: “The government I did in Bogotá reduced poverty by a third, closed the quality gap between private and public education, brought infant mortality to a single digit and hunger to zero. It delivered 45 schools, 3 university campuses, 1,200 preschool classrooms, opened the San Juan.”[9]

Example of Duque’s tweet: “If Colombia bets on the sectors of the Orange Economy, I’m sure that our country will connect digitally and we will have an opportunity for this economy to grow and expand the middle class. #The Future Of All pic.twitter.com/DWXSwkK9oc”.[10]

[6] https://twitter.com/ivanduque/status/1001518660375187456
[7] https://twitter.com/petrogustavo/status/978097683251507200
[8] https://twitter.com/IvanDuque/status/994694071913537536
[9] https://twitter.com/petrogustavo/status/1002973896579141633
[10] https://twitter.com/IvanDuque/status/1002742950273699842

Table 2. Summary of the Similarities and Differences Uncovered Automatically

Score	Duque	Petro
Higher	Education, Work, Violence, Economy, Environment, and Housing	Energy, Poverty, and Rights
Similar	**Housing:** ‘access’ to ‘decent’ and ‘own’ housing. Rights-based approach. **Violence:** direct allusion, including gender-based violence. **Energy:** traditional terminology and state management. **Poverty:** bureaucratic and local. The family as a victim.	**Housing:** terms do not clearly indicate a semantic axis. **Violence:** metaphorical and more intense allusion. More ideological. **Energy:** qualifying or taxonomic adjectives, ecological lexicon and technical knowledge **Poverty:** delimited, active and resolutive character.
Different	**Health:** system management	**Health:** reference to the patient

6.3 Categories with Similar Scores among the Candidates

There are three categories where the lexical pertinence scores of both candidates are very close: Health, Corruption, and Youth. However, they align with high pertinence in Health and low pertinence in Corruption and Youth.

In the Health category, Duque’s most pertinent terms are: ‘education’, ‘prevention’, ‘problems’, ‘quality’, ‘hospital’, ‘it’, ‘service’, ‘sustainable’, ‘system’ and ‘sustainability’. Petro’s top terms are: ‘prevention’, ‘treatment’, ‘medical’, ‘need’, ‘coverage’, ‘economic’, ‘patient’, ‘strengthening’, ‘preventive’ and ‘achieve’. Duque’s lexicon emphasizes the management and quality of the health system as a whole, while Petro’s vocabulary focuses on specific uses of the system, such as prevention, treatment, and patient care.

Although both candidates are equally pertinent to this topic, comparing their lexicon reveals semantic differences in their priorities and fields of application, reflecting differing ideological positions. Such nuanced understanding cannot be fully captured by automated data recording alone. While algorithms can automate certain aspects of semantic analysis, a comprehensive understanding of language use requires expert interpretation within context.

Table 2 summarizes the similarities and differences instinctively found between the discourses of the two political leaders.

7 Discussion

The application of our algorithms has enabled the detection of the most relevant terms used by the two main candidates in Colombia’s 2018 presidential elections, Iván Duque and Gustavo Petro, concerning key political agenda issues important to the Latin American population. These terms were extracted from a sample of 6,139 tweets from their official X accounts (@IvanDuque and @petrogustavo) posted between January 1 and June 17, 2018, marking the end of the electoral campaign. The pertinence of the selected words was validated through application to a Wikipedia corpus and verified by 300 undergraduate and graduate communication students.

Our results validate the model’s ability to operationalize the concept of linguistic pertinence. Pertinence emerges as a significant feature for differentiating the lexical choices of each politician in relation to a shared agenda, representing an initial level of expression of the “interpretative models of social reality” from which they derive meaning for their words. The production of words aligned with shared agendas is a consistent feature of political language, and discursive differentiation, especially during electoral competition, highlights ideological differences. Therefore, variations in pertinence can indicate differences in ideological positioning.

We observed that Duque and Petro differ in the subjects where they exhibit higher pertinence. The choice of a more or less pertinent lexicon for each topic reflects the importance they assign to the topic and their distinct approaches. The most relevant language is typically the most concrete, pragmatic, and direct, making it easier to infer the category addressed without explicit mention, through internal semantic analysis. The algorithm's ability to link terms to categories surpasses keyword occurrence analysis by retrieving words relevant to the subject even when it is not explicitly mentioned.

Our findings confirm that the categories where Duque and Petro's words are most pertinent align with their campaign proposals, which are the most programmatic and technical definitions, such as Education for Duque and Energy for Petro. Duque employs the language of state management when discussing Education or Health, while Petro more precisely delineates the problem of Poverty and adopts a proactive stance. Conversely, in some categories, both politicians, especially Petro, choose less direct, connotative, or metaphorical terms. Interpreting these metaphors requires contextual analysis, considering the complete tweet and the entirety of their tweets.

The model's ability to distinguish between the candidates' lexical choices and link them to broader ideological frameworks offers a valuable tool for political discourse analysis. It highlights the significance of language as a strategic element in political communication, reflecting deeper ideological commitments and policy priorities. Future research could expand this approach to other contexts and political figures, further exploring the intricate relationship between language, ideology, and political strategy.

Moreover, this study provides a foundational framework for employing computational linguistic models in political discourse analysis. The integration of human judgment to validate algorithmic outputs bridges the gap between quantitative and qualitative research methods, offering a comprehensive understanding of political rhetoric. This hybrid approach can be particularly beneficial in educational settings, where students can learn to critically analyze political language using both computational tools and traditional linguistic methods.

In addition, the relevance scores generated by our model can serve as predictive indicators of political success. By analyzing the alignment between a candidate's language and public concerns, political analysts and strategists can tailor campaign messages to resonate more deeply with voters. This application underscores the practical implications of our research, extending beyond academic inquiry to real-world political strategy and public engagement.

Finally, our research underscores the importance of transparency and accountability in political communication. By quantifying the pertinence of political discourse, we can hold candidates accountable for their rhetorical commitments and ensure that their language reflects genuine engagement with public issues. This methodological approach could be adapted for use by media organizations, watchdog groups, and civic educators to promote a more informed and active electorate.

The case analyzed allows us to propose a model to measure the relevance of a politician's lexical choices in a set of categories. We believe that this model can be applied to different cases of political discourse, and we present a procedure for categorizing the words to be analyzed and an algorithm for assigning the speakers' lexical choices to these categories. The ability to reproduce the context of occurrence of words is an indicator of relevance at the textual level. In addition, the hierarchization of the themes and the intensity of the selected words is a first ideological indicator. At this level, we tested the hypothesis that a list of words is more ideologically loaded when its denotation is less direct and its connotation is more intense.

References

[1] Tatiana Ananko. 2017. The category of evaluation in political discourse. *Advanced Education* 4, 8 (2017), 128–137.

[2] Natalia Aruguete. 2016. El poder de la agenda. (2016).

[3] Natalia Aruguete and Mario Riorda. 2014. ¿ Ideología u homogeneización? un análisis de las campañas electorales de Argentina, México y Venezuela. *Revista Mexicana de Opinión Pública* 16 (2014), 30–49.

[4] Aitolkyn Baigutanova, Jaehyeon Myung, Diego Saez-Trumper, Ai-Jou Chou, Miriam Redi, Changwook Jung, and Meeyoung Cha. 2023. Longitudinal assessment of reference quality on Wikipedia. In *Proceedings of the ACM Web Conference 2023*. 2831–2839.

[5] Jaime Durán Barba. 2017. *La política en el siglo XXI: Arte, mito o ciencia*. Debate.

[6] Jaime Durán Barba and Santiago Nieto. 2011. *El arte de ganar: Cómo usar el Ataque en campañas Electorales Exitosas*. Debate.

[7] Pablo Barberá and Gonzalo Rivero. 2015. Understanding the political representativeness of Twitter users. *Social Science Computer Review* 33, 6 (2015), 712–729. https://doi.org/10.1177/0894439314558836

[8] Francesco Barbieri, Horacio Saggion, and Francesco Ronzano. 2014. Modelling sarcasm in Twitter, a novel approach. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, Baltimore, Maryland, 50–58.

[9] Larry M. Bartels. 1996. Uninformed votes: Information effects in presidential elections. *American Journal of Political Science* 40, 1 (1996), 194–230. http://www.jstor.org/stable/2111700

[10] Jacopo Bindi, Davide Colombi, Flavio Iannelli, Nicola Politi, Michele Sugarelli, Raffaele Tavarone, and Enrico Ubaldi. 2018. Political discussion and leanings on Twitter: The 2016 Italian constitutional referendum. *arXiv preprint arXiv:1805.07388* null, null (2018), 24 pages.

[11] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, null (2003), 993–1022.

[12] Sergio Boisier. 2013. *The Origins, Development and Current State of Territorial Policies in Latin America in the Twentieth and Twenty-First Centuries*. Vol. 84. Springer International Publishing, Berlin, 69–89. https://doi.org/10.1007/978-3-642-39674-8_4

[13] Igor Brigadir, Derek Greene, and Pádraig Cunningham. 2015. Analyzing discourse communities with distributional semantic models. In *Proceedings of the ACM Web Science Conference*. Association for Computing Machinery, New York, NY, USA, Article 27, 10 pages.

[14] James Bryce. 2002. *Political Organizations in the United States and England, 1893*. 239–244. https://doi.org/10.1057/9780230107403_33

[15] Rodrigo Castro and Carmen Vaca. 2017. National leaders' Twitter speech to infer political leaning and election results in 2015 Venezuelan Parliamentary Elections. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW'17)*. IEEE, New Orleans, LA, USA, 866–871.

[16] Seungjin Choi. 2008. Algorithms for orthogonal nonnegative matrix factorization. *Proceedings of the International Joint Conference on Neural Networks* 1, 1 (2008), 1828–1832.

[17] Mathieu Cliche. 2017. BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. Association for Computational Linguistics, Vancouver, Canada, 573–580.

[18] I. Crespo, O. D'Adamo, V. García, and A. Mora. 2016. *Diccionario Enciclopédico de Comunicación Política*. Centro de Estudios Políticos y Constitucionales, Madrid. http://www.alice-comunicacionpolitica.com/wikialice

[19] Robert A. Dahl. 2004. Los sistemas políticos democráticos en los países avanzados: éxitos y desafíos. *CLACSO, Consejo Latinoamericano de Ciencias Sociales Editorial* (2004), 24–36.

[20] Marina Díaz-Peralta. 2018. Metaphor and ideology: Conceptual structure and conceptual content in Spanish political discourse. *Discourse and Communication* 12, 2 (2018), 128–148. https://doi.org/10.1177/1750481317745752

[21] Umberto Eco, Rorty Richard, Culler Jonathan, and Christine Brooke-Rose. 1992. *Interpretation and Overinterpretation*. Cambridge University Press, Cambridge.

[22] Saad Ezzini, Sallam Abualhaija, and Mehrdad Sabetzadeh. 2022. Wikidominer: Wikipedia domain-specific miner. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 1706–1710.

[23] Ilias Flaounas, Marco Turchi, Omar Ali, Nick Fyson, Tijl De Bie, Nick Mosdell, Justin Lewis, and Nello Cristianini. 2010. The structure of the EU mediasphere. *PloS One* 5, 12 (2010), e14243.

[24] Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'18)*. European Language Resources Association (ELRA), Miyazaki, Japan, 3483–3487. https://aclanthology.org/L18-1550

[25] A. J. Greimas. 1987. *Semántica Estructural: Investigación Metodológica*. Gredos, Madrid.

[26] Justin Grimmer and Brandon M. Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21, 3 (2013), 267–297.

[27] David W. Johnson and Roger T. Johnson. 2000. Civil political discourse in a democracy: The contribution of psychology. *Peace and Conflict: Journal of Peace Psychology* 6, 4 (2000), 291–317. https://doi.org/10.1207/s15327949pac0604_01

[28] Tom Kenter and Maarten de Rijke. 2015. Short text similarity with word embeddings. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (Melbourne, Australia) *(CIKM '15)*. Association for Computing Machinery, New York, NY, USA, 1411–1420.

[29] Catherine Kerbrat-Orecchioni, Sara Vassallo, and Eduardo Villamil. 1983. *La Connotación*. Hachette, Buenos Aires.

[30] Sopan Khosla, Niyati Chhaya, and Kushal Chawla. 2018. Aff2Vec: Affect–enriched distributional word representations. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2204–2218.

[31] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* (2014). 1746–1751. https://doi.org/10.3115/v1/d14-1181

[32] Donald R. Kinder and Thomas R. Palfrey. 1993. On behalf of an experimental political science. *Experimental Foundations of Political Science* (1993), 1–39.

[33] G. Lakoff. 2016. Moral Politics: How Liberals and Conservatives Think, Third Edition. University of Chicago Press, "Chicago".

[34] G. Lakoff and M. Johnson. 1986. Metáforas de la vida cotidiana. Cátedra.

[35] Walter Lippmann. 1922. Public Opinion. Brace & Co.

[36] Yang Liu, Alan Medlar, and Dorota Glowacka. 2021. Statistically significant detection of semantic shifts using contextual word embeddings. *arXiv preprint arXiv:2104.03776* (2021).

[37] Maxwell McCombs and Sebastián Valenzuela. 2007. The agenda-setting theory. *Cuadernos de Información* 20 (2007), 44–50.

[38] Maxwell E. McCombs and Donald L. Shaw. 1972. The agenda-setting function of mass media. *Public Opinion Quarterly* 36, 2 (1972), 176–187.

[39] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Atlanta, Georgia, 746–751. https://aclanthology.org/N13-1090

[40] Pedro Miguel Moas and Carla Teixeira Lopes. 2023. Automatic quality assessment of Wikipedia articles–a systematic literature review. *Comput. Surveys* 56, 4 (2023), 1–37.

[41] Elin Naurin, Stuart Soroka, and Niels Markwat. 2019. Asymmetric accountability: An experimental investigation of biases in evaluations of governments' election pledges. *Comparative Political Studies* 52, 13-14 (2019), 2207–2234. https://doi.org/10.1177/0010414019830740

[42] E. Núñez. 2000. *Aproximación al Léxico del Lenguaje Político Español*. Ph. D. Dissertation. Tesis doctoral). Universidad de Málaga, España.

[43] María Soledad Padilla Herrada. 2015. La argumentación política en Twitter. *Discurso & Sociedad* 9, 4 (2015), 419–444.

[44] Sebastian Padó, Andre Blessing, Nico Blokker, Erenay Dayanik, Sebastian Haunss, and Jonas Kuhn. 2019. Who sides with whom? Towards computational construction of discourse networks for political debates. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 2841–2847. https://doi.org/10.18653/v1/P19-1273

[45] Ana Pano Alamán and Ana Mancera Rueda. 2013. Nuevas dinámicas discursivas en la comunicación política en Twitter. *Círculo de Lingüística Aplicada a la Comunicación* null, 56 (2013), 53–80.

[46] Gregory Park, H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Michal Kosinski, David J. Stillwell, Lyle H. Ungar, and Martin E. P. Seligman. 2015. Automatic personality assessment through social media language. *Journal of Personality and Social Psychology* 108, 6 (2015), 934.

[47] Marco Pennacchiotti and Ana-Maria Popescu. 2011. A machine learning approach to Twitter user classification. In *Fifth International AAAI Conference on Weblogs and Social Media*, Vol. 5. AAAI Press, Palo Alto, California USA, 281–288.

[48] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543.

[49] Oswaldo Pita, Gabriela Baquerizo, Carmen Vaca, Jonathan Mendieta, Mónica Villavicencio, and Jorge Rodríguez. 2016. Linguistic profiles on microblogging platforms to characterize political leaders: the Ecuadorian case on Twitter. In *2016 IEEE Ecuador Technical Chapters Meeting (ETCM'16)*. IEEE, null, 1–6.

[50] Adithya Rao and Nemanja Spasojevic. 2016. Actionable and Political Text Classification using Word Embeddings and LSTM.

[51] B. Reinalda. 2013. *Routledge Handbook of International Organization*. Taylor & Francis, null.

[52] Ludovic Rheault and Christopher Cochrane. 2019. Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis* 28, 1 (2019), 112–133. https://doi.org/10.1017/pan.2019.26

[53] Mariano Rojas. 2019. *Well-Being in Latin America: Drivers and Policies*. Springer International Publishing, Switzerland. 1–183 pages.

[54] Md. Safiullah, Pramod Pathak, Saumya Singh, and Ankita Anshul. 2017. Social media as an upcoming tool for political marketing effectiveness. *Asia Pacific Management Review* 22, 1 (2017), 10–15. https://doi.org/10.1016/j.apmrv.2016.10.007

[55] Joseph A. Schumpeter. 2013. *Capitalism, Socialism and Democracy*. Routledge.

[56] Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2018. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political Communication* 35, 1 (2018), 50–74. https://doi.org/10.1080/10584609.2017.1334728

[57] Eric S. Tellez, Sabino Miranda-Jiménez, Mario Graff, Daniela Moctezuma, Oscar S. Siordia, and Elio A. Villaseñor. 2017. A case study of Spanish text transformations for Twitter sentiment analysis. *Expert Systems with Applications* 81 (2017), 457–471.

[58] Robert Thomson and Heinz Brandenburg. 2019. Trust and citizens' evaluations of promise keeping by governing parties. *Political Studies* 67, 1 (2019), 249–266. https://doi.org/10.1177/0032321718764177

[59] Teun A. Van Dijk. 2005. Discourse analysis as ideology analysis. In *Language & peace*. Routledge. 41–58.

[60] Teun A. van Dijk. 2008. Text, context and knowledge. *Universitat Pompeu Fabra* (2008), 38.

[61] Aniela Suray Ventura. 2016. Argumentación y discurso político en Twitter. Análisis de la campaña presidencial argentina 2011. *Aposta. Revista de Ciencias Sociales* 69 (2016), 39–66.

[62] Eliseo Verón. 1995. Semiosis de lo ideológico y del poder. *Espacios de Crítica y Producción* 1 (1995), 43–51.

Received 6 September 2023; revised 5 August 2024; accepted 9 August 2024