Text Analysis for Information Extraction in Political Discourse



Abdelouahab BERREHAIL

University of Batna 1, Algeria, Abdelouahab.berrehail@univ-batna.dz

Received date: 24/04/2026 Accepted date: 17/05/2026 Publication date: 01/07/2026

Abstract:

This study investigates the application of text analysis techniques for information extraction in political discourse. Due to the growing volume and complexity of political texts, traditional analytical methods are often insufficient to fully capture their meaning and structure. The research explores how natural language processing and artificial intelligence can be used to extract structured and meaningful information from political discourse. The study focuses on key techniques such as Named Entity Recognition, sentiment analysis, and topic modeling to identify political actors, evaluate attitudes, and detect dominant themes. The findings indicate that these methods enhance the accuracy and efficiency of political text analysis, despite challenges related to ambiguity and context. The study concludes that combining computational tools with discourse analysis provides deeper insights into political communication.

Keywords: *political discourse; text analysis; information extraction; NLP; artificial intelligence.*

*** Corresponding author: Abdelouahab BERREHAIL, Abdelouahab.berrehail@univ-batna.dz**

Introduction:

In recent decades, the expansion of digital communication and political content has increased the importance of analyzing political discourse. Text analysis, particularly for information extraction, has become a key methodological approach for understanding the structure, meaning, and intentions of political texts. By using computational and linguistic techniques, researchers can identify themes, patterns, and entities within large datasets.

Beyond simple interpretation, political discourse analysis seeks to uncover implicit messages, ideological positions, and strategic narratives. Techniques such as named entity recognition, sentiment analysis, and topic modeling enable the transformation of unstructured text into structured data. The integration of natural language processing and machine learning further enhances the accuracy and efficiency of analysis, contributing to a deeper understanding of political communication and behavior.

Based on the above, the research problem can be formulated in the form of the following main question:

How can text analysis techniques be employed to extract accurate and structured information from political discourse?

This main question gives rise to a set of sub-questions:

- What are the linguistic and semantic features that characterize political discourse and influence the process of information extraction?
- What are the main text analysis, artificial intelligence, and natural language processing techniques used in information extraction?

The study hypothesizes that the use of text analysis techniques based on artificial intelligence and natural language processing enables the extraction of accurate and structured information from political discourse, despite its complexity and ambiguity.

To address this research problem, the study adopts a descriptive-analytical methodology.

1. Definition:

Text analysis for information extraction in political discourse refers to a set of methods derived from Natural Language Processing that aim to automatically identify, structure, and interpret relevant information—such as entities, relationships, topics, and opinions—within political texts, in order to facilitate their analytical understanding and use in political and social research.

According to Daniel Jurafsky and James H. Martin ; nformation extraction refers to the process of transforming unstructured text into structured data, including entities, relations, and events (Jurafsky & Martin, 2026, p. 455)

2. Specificities of Political Discourse:

Political discourse is characterized by a set of linguistic, rhetorical, and pragmatic features that distinguish it from other types of discourse. It is not merely a neutral vehicle of information but a strategic form of communication aimed at persuasion, legitimation, and influence within a given sociopolitical context.

a. Persuasive and Strategic Nature:

One of the primary characteristics of political discourse is its inherently persuasive nature. Political actors use language strategically to influence public opinion, justify policies, and gain support. This persuasion is often achieved through argumentation, emotional appeal, and selective presentation of information.

Political discourse is fundamentally oriented toward persuasion and the exercise of power through language (Chilton, 2004, p. 4). Similarly, Teun A. van Dijk emphasizes that political discourse functions as a form of social action aimed at shaping cognition and public attitudes. In this context, political discourse is not limited to the actions of politicians alone. As he argues, although politicians are crucial actors and authors of political discourse and other political practices, they are not the only participants in the political domain. From an interactional perspective, discourse analysis must also take into account the various recipients involved in political communication, including the public, citizens, and other social groups. Once political discourse is situated within the public sphere, a wider range of participants emerges, highlighting the complex and interactive nature of political communication (van Dijk, 1997, pp. 12-13).

b. Ideological and Interpretative Dimension:

Political discourse is deeply embedded in ideology. It reflects and reproduces systems of beliefs, values, and power relations. Language is used not only to describe reality but also to construct and frame it according to specific ideological perspectives.

According to Norman Fairclough, discourse plays a central role in the construction and dissemination of ideology, shaping how social realities are perceived (Fairclough, 1995, p. 14). This interpretative dimension makes political discourse a key object of critical analysis.

c. Use of Rhetorical and Figurative Devices:

Political discourse frequently relies on a range of rhetorical strategies—such as metaphors, analogies, repetition, and slogans—to render complex political issues more accessible and persuasive. These devices do not merely simplify information; they actively structure how audiences interpret reality by highlighting certain aspects while downplaying others. Repetition and slogans, for instance, reinforce key messages and enhance their memorability, while analogies help

bridge unfamiliar political concepts with everyday experiences, thereby facilitating comprehension and emotional engagement (Charteris-Black, 2011, p. 28).

d. Ambiguity and Strategic Vagueness:

Another defining feature of political discourse is the strategic use of ambiguity and vagueness. Political actors frequently employ deliberately indeterminate language to appeal to heterogeneous audiences, avoid firm commitments, and preserve interpretive flexibility. Rather than being a weakness, such vagueness functions as a calculated communicative resource that enables speakers to navigate complex political environments. By leaving room for multiple interpretations, politicians can simultaneously address conflicting interests and sustain broader support bases. In this sense, ambiguity operates as a pragmatic strategy that enhances the adaptability of political communication and broadens its persuasive reach (Chilton, 2004, p. 45).

e. Context-Dependence and Intertextuality:

Political discourse is inherently context-dependent, as its meaning is shaped by shared knowledge, historical references, and the broader socio-political environment in which it is produced and interpreted. It is also fundamentally intertextual, drawing upon and responding to prior texts, speeches, media narratives, and political events. Through this intertextual layering, political messages gain depth and legitimacy, as they resonate with familiar ideas and previously established discursive frames. In this sense, intertextuality links individual texts to wider social and political contexts, making political discourse dynamic and continuously evolving (Fairclough, 1995, p. 101).

3. Main Methods: Computational and Linguistic Approaches to Information Extraction in Political Discourse

Text analysis for information extraction in political discourse relies on a set of well-established computational and linguistic methods. These approaches aim to transform unstructured textual data into structured information that can be systematically analyzed. In this context, the main methods encompass a range of techniques designed to capture different dimensions of meaning and structure within political texts. The most prominent among these include Named Entity Recognition, Sentiment Analysis, Topic Modeling, and Syntactic Parsing, each contributing to a deeper and more systematic understanding of political discourse.

a. Named Entity Recognition (NER):

Named Entity Recognition (NER) is a fundamental technique in natural language processing used to identify and classify key elements in text, such as names of persons, organizations, locations, and political institutions. In the context of political discourse, NER enables researchers to detect the main actors and institutions involved in a narrative, thereby facilitating network and actor-based

analysis. As a core component of information extraction systems, NER contributes to structuring unstructured textual data into clearly identifiable units. Moreover, by linking textual elements to real-world referents, it enhances interpretability and supports more precise analytical outcomes (Kale & Govilkar, 2017, p. 5).

b. Sentiment Analysis:

Sentiment analysis focuses on detecting the emotional tone or evaluative stance expressed in a text. In political discourse, this method is widely used to assess public opinion, political attitudes, and rhetorical strategies. It enables the classification of textual data into categories such as positive, negative, or neutral, and in more advanced applications, it can capture more nuanced emotional states and evaluative judgments. Within this framework, sentiment analysis is defined as the computational study of opinions, sentiments, and emotions expressed in text (Liu, 2012, pp. 10–13). This approach is particularly valuable in political contexts, as it allows researchers to analyze speeches, media content, and social discourse in order to uncover persuasive strategies, ideological positioning, and audience targeting.

c. Topic Modeling:

Topic modeling is an unsupervised machine learning technique used to uncover latent themes within large corpora of text. Among the most widely applied models is Latent Dirichlet Allocation (LDA), which identifies clusters of co-occurring words that represent underlying topics.

This method allows for the automatic discovery of hidden thematic structures in large collections of documents without the need for prior annotation (Blei et al., 2003, p. 993). In the context of political discourse analysis, topic modeling is particularly valuable for revealing dominant issues, identifying ideological trends, and highlighting policy priorities.

d. Syntactic and Dependency Parsing:

Syntactic parsing is concerned with analyzing the grammatical structure of sentences, whereas dependency parsing focuses on identifying the relationships between words within a sentence. Together, these approaches are essential for understanding how meaning is constructed in political discourse, particularly in complex or ambiguous statements.

Syntactic analysis provides deeper insight into sentence structure and supports more accurate semantic interpretation, enabling a more precise extraction of meaning (Jurafsky & Martin, 2026, p. 457). This is especially relevant in political texts, where linguistic structures often convey implicit, strategic, or persuasive meanings.

e. Information Extraction Systems:

Information Extraction (IE) systems aim to transform unstructured textual data into structured representations by identifying relevant entities, relationships, and other useful information. These systems support information retrieval and text analysis tasks by enabling the organization and indexing of textual content in a structured form (Manning et al., 2008, pp. 249–250).

4. Information Extraction Techniques

Information extraction (IE) refers to the process of automatically extracting structured information from unstructured textual data. In the context of political discourse, IE techniques enable the identification of entities, relationships, events, and semantic patterns, thereby transforming complex texts into analyzable data.

a. Machine Learning-Based Techniques:

Machine learning approaches rely on annotated datasets to train models capable of recognizing patterns and extracting relevant information. Supervised learning methods, in particular, learn from labeled examples to make predictions on new data (Bishop, 2006, p. 3).

These approaches enable systems to generalize from training data to unseen instances by learning underlying statistical regularities. They are widely applied in natural language processing tasks such as text classification, named entity recognition, and relation extraction, where accurate pattern recognition is essential for transforming unstructured text into meaningful and structured information.

b. Deep Learning Approaches:

Deep learning has revolutionized information extraction by enabling models to automatically learn complex representations of language. Techniques such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Transformers are widely used in modern IE systems.

According to Jacob Devlin et al. (2019), transformer-based models like BERT achieve state-of-the-art performance in multiple NLP tasks, including information extraction (Devlin et al., 2019, pp. 4171–4172). These models are particularly effective in capturing contextual meaning in political texts.

Conclusion:

In summary, text analysis techniques offer powerful tools for extracting structured information from political discourse. By leveraging natural language processing and artificial intelligence, researchers can better identify key actors, themes, and underlying meanings within complex texts. Despite challenges related to ambiguity and context, these methods significantly enhance the efficiency and accuracy of analysis. Combining computational approaches with qualitative interpretation remains essential for deeper insights. Ultimately, this approach

contributes to a more systematic and informed understanding of political communication.

Bibliography:

- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York: Springer.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.
- Charteris-Black, J. (2011). Politicians and rhetoric: The persuasive power of metaphor (2nd ed.). Palgrave Macmillan.
- Chilton, P. (2004). Analysing political discourse: Theory and practice. Routledge.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186.
- Fairclough, N. (1995). Critical discourse analysis: The critical study of language. Longman.
- Jurafsky, D., & Martin, J. H. (2026). Speech and Language Processing (3rd ed., draft). Stanford University.
- Kale, S., & Govilkar, S. (2017). Survey of named entity recognition techniques for various Indian regional languages. International Journal of Computer Applications, 164(4), 37–43.
- Liu, B. (2012). Sentiment analysis and opinion mining. Morgan & Claypool Publishers. https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.
- van Dijk, T. A. (1997). What is political discourse analysis? In J. Blommaert & C. Bulcaen (Eds.), Political linguistics (pp. 11–52). John Benjamins.