



## DATIS

### Data for Inclusive Societies: Foes and Friends of Inclusiveness in contemporary Greece

#### D3.2. Report on the text analysis (automated and coding)

**School of Political Sciences, Aristotle University of Thessaloniki**

**October 2025**



**H.F.R.I.**  
Hellenic Foundation for  
Research & Innovation

**Greece 2.0**  
NATIONAL RECOVERY AND RESILIENCE PLAN



Funded by the  
European Union  
NextGenerationEU

## Introduction to manifesto coding

Party manifestos represent the official party policy platforms and set the agenda for electoral contests. In the framework of the DATIS project, one of the main objectives was to gather and systematically code party manifestos from every election that occurred during the project. Within this timeframe, the only relevant electoral event was the 2024 European Parliament election, which therefore became the primary focus of our data collection and analysis efforts. DATIS concentrates on the study of social groups, with the overarching goal of enabling systematic analyses of political parties' positions on inclusive societies. Because party manifestos outline not only parties' views on social inclusion but also on a wide range of additional policy domains, it is crucial to preserve the full breadth of this information. By coding every quasi-sentence of all available party manifestos using the Euromanifesto Coding Scheme, this deliverable provides a dataset well suited for analyses of broader political interest. Researchers can draw on this data set to study, among others, how parties conceptualize social groups, allocate attention to inclusive policies, and embed inclusion-related issues within broader political agendas, allowing for a more comprehensive understanding of the political foundations of inclusive societies.

**Figure 1** illustrates the added value that our data bring to the study of inclusive societies in Greece, by plotting the differences in social group appeals between included parties. Further the use of the Euromanifesto coding scheme, paired with the ParlGov and Euromanifesto party codes, makes these data interoperable, and easy to merge with prior editions of the Euromanifesto project.

## Social Group Salience in EP 2024 Manifestos

Counts of quasi-sentences by party and issue category

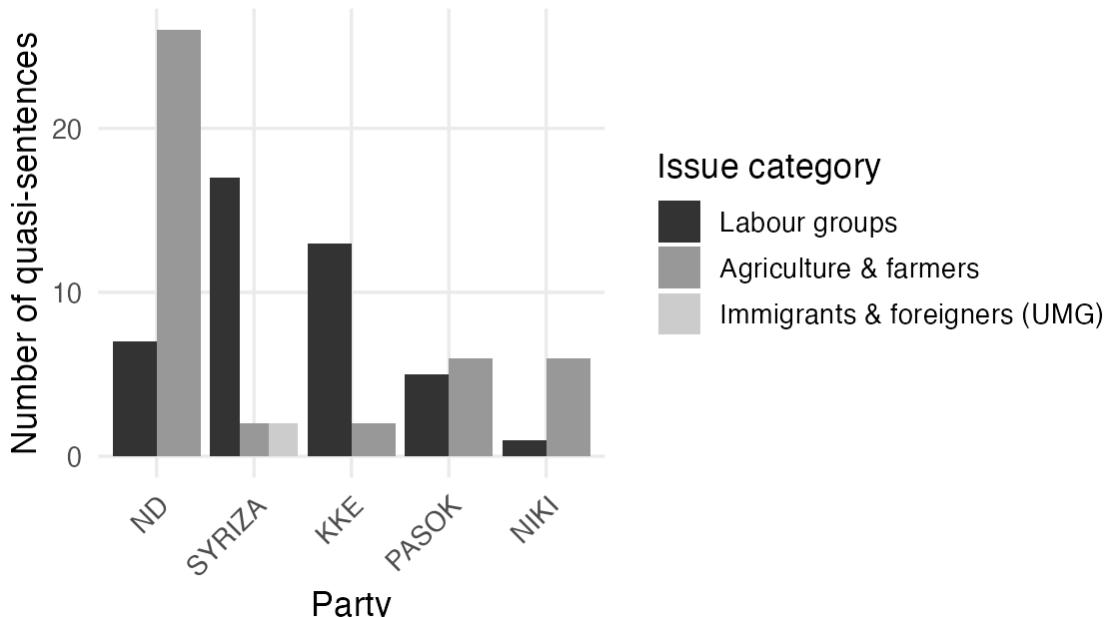


Figure 1 The Relative Prevalence of Social Group Codes in the Coded Manifestos

## Scope, Sources, and Methodology of Manifesto Coding

This study set out to code manifestos for all elections in the span of the project. This includes only the European Parliament (EP) elections of 2024. For this reason, we have applied the Euromanifesto (EMCS V) coding scheme (Braun & Reinl, 2022)<sup>1</sup>. In terms of party selection, we have included all parties that secured representation in the European Parliament, with the exception of Elliniki Lisi (EL), for which we could not identify any published party manifesto or functionally equivalent document. Consequently, the data are comprised of the following parties: New Democracy (ND), Synaspismos Rizospastikis Aristeras (SYRIZA), Panhellenic Socialist Movement (PASOK), Democratic Patriotic Popular Movement (NIKI), Kommounistiko Komma Ellados (KKE) and Foni Logikis (Voice of Reason).

The data include several party identifiers. The **party** variable denotes party acronyms as elaborated above. Further **emcs** and **parlgov\_id**, correspond to party codes, where available, as codified in the Euromanifesto and Parlgov projects, accordingly. These overlapping party

<sup>1</sup> For a crosswalk between Euromanifesto and Manifesto Project codes see [https://manifestoproject.wzb.eu/information/documents/cmp\\_emp\\_mapping](https://manifestoproject.wzb.eu/information/documents/cmp_emp_mapping)

identification variables aim to make the merging with prior Euromanifesto data and auxiliary information easier for researchers.

Both the coded and uncoded data sets (see D3.3) retain the original quasi sentences (in Greek) in the **quasi\_sentence** variable, and the **type\_of\_manifesto** variable indicates the specific kind of source material coded (full manifesto/policy pamphlet/ party leader statement), while **source** provides the origin link. Further, per Euromanifesto conventions, the **level** variable indicates whether each quasi sentence refers to the national, European, or global level. The **direction** variable (positive/negative) captures quasi-statements valence. Finally, the **domain** variable acts as a conceptual organizing variable, denoting the domain each **code** belongs to, per the Euromanifesto scheme. The party family classification (**party\_family**) used in this dataset originates from the expert assessment of the research team.

Table 1 Type of Manifesto Code, by Party

Party	Type of Manifesto
ND	Policy Pamphlet
NIKI	Full Manifesto
PASOK	Full Manifesto
SYRIZA	Full Manifesto
Voice of Reason	Party Leader Statement
KKE	Full Manifesto

Following the Euromanifesto workflow, the coding procedure proceeded in three stages (for an overview of the methodology see Braun & Reinal 2022). First, *quantification* involved determining the number of statements contained in each manifesto, and manually unitizing the text, accordingly. Main titles, section headings, and subsection titles were also included, as they often contain substantive programmatic cues relevant to understanding a party's issue emphases. Second, during *classification*, we assigned each quasi-sentence the appropriate issue code. Third, with *evaluation* we considered the manifesto as a whole to capture its broader thematic and situate codes in their appropriate context.

The EM Project treats *quasi-sentences* as the basic coding unit, defined as an explicit expression of a political idea or argument (Volkens et al., 2014). Punctuation and grammatical

structure generally guided the identification of these units, as each sentence typically conveys a distinct meaning. To ensure coding quality, reliability was assessed through systematic cross-checking against external benchmark texts and coder comparison procedures commonly used in manifesto research (Volkens et al., 2014).

## Social Media Data Collection

We employed a comprehensive data collection strategy focused on Greek political actors. We systematically downloaded all available public content from Facebook pages belonging to politicians from recognized Greek parties, covering a longitudinal period extending up to the end of September 2025. The resulting dataset is substantial, comprising approximately 400,000 posts generated by 166 distinct political actors. These political actors include party leaders, members of parliament, ministers, former ministers, and elected representatives with an active online presence. This offers a rich, extensive, and statistically significant corpus for observing the evolution of political communication and enables a deep-dive analysis into public sentiment analysis and the discursive construction of inclusiveness, providing evidence-based conclusions on the health of digital democracy in Greece.

Data were retrieved either from public Facebook pages or, where applicable, from public personal profiles, provided that they met Facebook's accessibility criteria for research use. Specifically:

- **Public Pages:** No restrictions were applied regarding follower count or interaction volume; all available posts were collected.
- **Personal Profiles:** Data were accessed only when the account satisfied Facebook's public visibility threshold of 10,000 followers or more, ensuring compliance with platform constraints and ethical standards.

All posts were scraped using automated collection tools that capture post text, timestamps, and publicly available engagement indicators (reactions, shares, and comments).

# Methodology

## Topic Modelling

To identify the main themes of political discourse, we applied topic modelling techniques on the cleaned corpus. Topic modeling, according to Robinson & Silge (2017), is a method for unsupervised classification of such documents, similar to clustering in numerical data, which finds natural groups of objects even when the researcher is not looking for a specific topic.

Latent Dirichlet Allocation (LDA) served as the primary modelling approach, enabling the detection of clusters of semantically coherent topics across the political spectrum. Latent Dirichlet Allocation (LDA) is a generative probabilistic model and one of the most prevalent algorithms for topic modeling. The methodology operates on the premise that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. Unlike discrete clustering methods, LDA allows for thematic "overlap," reflecting the polysemic nature of natural language where documents often address multiple subjects simultaneously.

The LDA framework is governed by two fundamental distributional principles:

- **Document-Topic Distribution:** Each document is modeled as a multinomial distribution of topics. This allows for a granular representation of content; for instance, a single document may be composed of Topic A and Topic B in varying proportions (e.g., a 90/10 or 30/70 split).
- **Topic-Word Distribution:** Each latent topic is modeled as a distribution over a fixed vocabulary. In a dual-topic model concerning national news, a "politics" topic would exhibit a high probability for terms such as *President* and *Congress*, whereas an "entertainment" topic would favor terms like *cinema* or *television*. Crucially, words are not mutually exclusive and may appear across multiple topics with varying densities.

Mathematically, LDA serves as a dual-estimation tool: it simultaneously identifies the word clusters that constitute specific topics and calculates the thematic proportions that define each document within a corpus.

Kang et al (2013) utilized topic modeling to conduct opinion mining on newspaper editorials, positing that media partisanship serves as a proxy for institutional opinion. The methodological

framework proceeded in four distinct stages. First, topics were extracted from a large-scale corpus of newspaper texts to determine their overall distribution. Second, the internal structure and thematic content of these topics were examined using network analysis. Third, a time-series analysis was employed to track the chronological distribution of topics across various publications. The findings indicated that both liberal and conservative newspapers exhibited reporting tendencies aligned with their respective ideologies. Consequently, the study demonstrated that topic-based opinion mining is a robust and reliable technique for analyzing latent perspectives within large textual datasets.

The Structural Topic Model (STM) represents a significant advancement in the field of machine-assisted reading, extending traditional latent variable models such as Latent Dirichlet Allocation (LDA). Developed by Roberts, Stewart, and Tingley(2019), the STM framework is specifically engineered for social science research, where the primary objective is often not just to discover themes, but to understand how these themes correlate with document-level characteristics. Unlike standard topic models that treat documents as exchangeable, the STM allows for the systematic incorporation of metadata—such as a politician’s party affiliation, gender, or the timing of a post—directly into the statistical estimation of the model (Roberts et al., 2014, 2016).

In the context of Greek political communication on Facebook, the STM is particularly advantageous. It enables the researcher to model two distinct types of relationships: **topical prevalence** and **topical content**. Topical prevalence refers to the proportion of a post dedicated to a specific subject (e.g., the economy vs. migration), which may vary according to the author’s ideological stance or the electoral cycle. Topical content, on the other hand, allows for the possibility that different politicians might use different vocabularies to discuss the same underlying topic. By leveraging the fast variational approximation implemented in the *stm* R package (Roberts et al., 2019), the model can process large-scale social media corpora while maintaining the interpretability required for political discourse analysis.

The implementation process begins with the ingestion and preprocessing of the Greek text corpus. Following standard text-as-data procedures (Grimmer et al., 2022; Grimmer & Stewart, 2013), the data is tokenized, and common stop words and infrequent terms are removed to reduce noise. A critical step in the STM workflow involves selecting the optimal number of topics (K). This is typically achieved through a multi-metric diagnostic process, evaluating the trade-off between semantic coherence—the degree to which words in a topic frequently co-

occur—and exclusivity, which measures the uniqueness of words to a specific topic. This ensures that the resulting topics are not only statistically robust but also substantively meaningful for the analysis of the Greek political landscape.

Finally, the STM allows for rigorous hypothesis testing through the estimation of covariate effects. By regressing topic proportions on metadata, researchers can quantify the influence of external factors on political messaging. For example, one can estimate how the salience of "populist" rhetoric differs between government and opposition parties over time. This approach transforms a purely unsupervised discovery process into a structured analytical tool, providing a reliable means of mapping the latent thematic structure of digital political communication (Roberts et al., 2019).

Topic distributions also enabled cross-party comparisons regarding the most salient issues emphasised by different political formations. Using the distributions of the topic, we also conducted a comparative analysis across parties. This enabled the identification of the issue priorities that differentiate parties from one another.

## **Sentiment Analysis**

Sentiment analysis, often referred to as opinion mining, is the computational study of opinions, sentiments, and emotions expressed within written text. As a primary subfield of Natural Language Processing (NLP), its applications have expanded beyond computer science into the management and social sciences, driven by the proliferation of digital discourse (Liu, 2012). The exponential growth of user-generated content has positioned sentiment analysis as a critical component of data mining, with feature extraction serving as its foundational step. Feature-based analysis enables a granular determination of user preferences by identifying specific entities or attributes (e.g., product features or political issues) and the sentiments associated with them. The advent of social media and web-based forums has provided an unprecedented volume of recorded human sentiment, which serves as a critical variable in understanding social behavior, political alignment, and organizational decision-making.

## **Dictionary-Based Methodologies and the "Bag-of-Words" Approach**

A common methodological framework for sentiment analysis involves a lexicographical approach, where a document is treated as a collection of individual words—a "bag-of-words"

model. In this framework, the sentiment of a text is calculated as the aggregate of the sentiment scores assigned to its constituent tokens. This approach is frequently implemented within the "tidy data" ecosystem, utilizing sentiment lexicons or dictionaries (Silge & Robinson, 2017).

These lexicons are typically constructed through crowdsourcing or expert annotation and validated against domain-specific data, such as product reviews or micro-blogging content. While effective, the application of these dictionaries requires caution when applied to corpora that differ significantly from the validation set, such as historical narratives or highly specialized technical texts. However, research suggests that sentiment can still be effectively measured by identifying shared terms between the lexicon and the target text, or through the use of domain-specific dictionaries tailored to particular fields (Silge & Robinson, 2017).

## **Methodological Limitations and Considerations**

Despite its utility, dictionary-based sentiment analysis possesses inherent limitations that must be addressed during data interpretation:

- **Contextual Qualifiers and Negation:** Simple aggregate methods often fail to account for linguistic qualifiers or negation (e.g., "not good"). While negation analysis can be performed by examining the frequency and proximity of negative tokens, the "bag-of-words" approach generally prioritizes individual word scores over complex syntactic structures.
- **Sarcasm and Linguistic Nuance:** These models often struggle with sarcasm or subtle emotional intent; however, in many large-scale corpora, these instances do not statistically outweigh the general sentiment trends.
- **Granularity of Analysis:** The scale of the textual unit—whether a sentence, paragraph, or entire document—significantly impacts the results. Larger text blocks tend to revert to a neutral mean as positive and negative sentiments aggregate, whereas analysis at the sentence or paragraph level typically yields higher precision and more distinct sentiment profiles.

Despite technical advancements, Clavel and Callejas (2016) highlight a divergence between the opinion mining and human-agent interaction (HAI) communities. While opinion mining focuses on large-scale sentiment detection, HAI requires real-time, socio-emotional processing. Their analysis suggests that current sentiment detection methods are often not

optimized for the time-constrained, interactive nature of human-agent dialogues. This underscores a growing need for interdisciplinary frameworks that adapt standard opinion mining techniques to facilitate more sophisticated socio-emotional interactions and strategic dialogue management.

## Findings

The initial dataset included a large volume of posts that consisted primarily of announcements regarding politicians' daily activities, event attendances, scheduled meetings, and ceremonial appearances. Although these posts play a functional role in political self-presentation and image management, they are not directly relevant to the analysis of policy-driven discourse.

The STM analysis generated a set of 61 topics, but we focus on dominant policy-related topics only. These topics were subsequently labelled, spanning areas such as economy, education, migration, social welfare, health, security, and human rights.

### Topic Prevalence

The analysis of topic prevalence shows that the highest weight in the corpus ( $\gamma$  coefficient) belongs to international affairs (~0.032), followed by issues of violence against women ("Women-Violence," ~0.018), sports (~0.017), health (~0.016), and education (~0.015). The associated keywords reveal the semantic composition of each topic. For example, the International topic includes words such as Greece, European, Europe, country, Turkey, Greek, while the Immigration topic is characterized by words like immigrant, country, migration, illegal, border, structure.

The prevalence hierarchy does not fully align with engagement patterns. International affairs are both the most frequently discussed and the most engaging for the audience, but other high-engagement topics, such as courts and banks, show lower prevalence. This suggests that even infrequently discussed topics can generate disproportionate engagement when addressed, indicating latent public demand for political discourse on these matters.

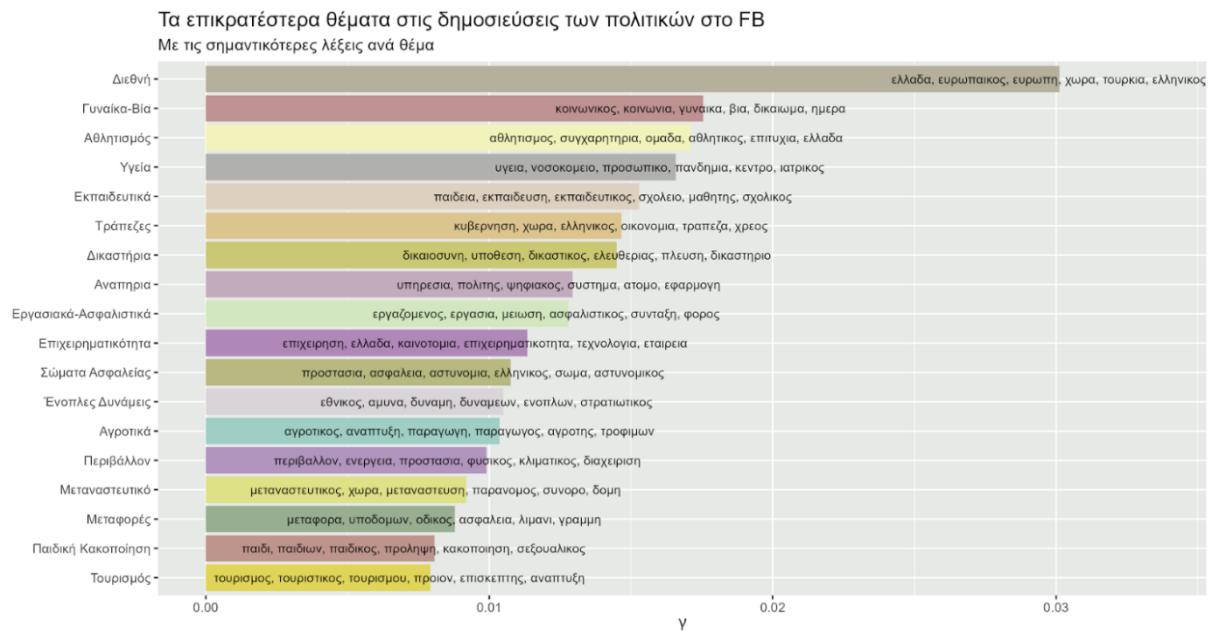


Figure 2 Top Topics in Political Posts on Facebook

### Comparison of Topic Prevalence between Government and Opposition Parties

Figure 3 visualizes the comparative standing of New Democracy (ND) and SYRIZA across various political and social topics, using topic modelling. The points to the right indicate the issues in which New Democracy has involved more frequently than SYRIZA, while points to the left indicate topics where SYRIZA is mostly focused on. More specifically, the figure visualizes the relative difference in thematic engagement between the two parties.

The distribution of topics reveals a clear thematic divide between the two parties. New Democracy focused more on economic and developmental sectors such as Agriculture, Tourism, Entrepreneurship, and Education, as well as on state security matters like Armed Forces and Security Forces. Conversely, SYRIZA maintains a strong lead in social and rights-based issues, with Labor and Insurance standing out as their strongest category, followed by Justice, Environment, and Migration. Moreover, the topic of Health hovers right on the zero line, suggesting that the difference between the two parties is negligible.

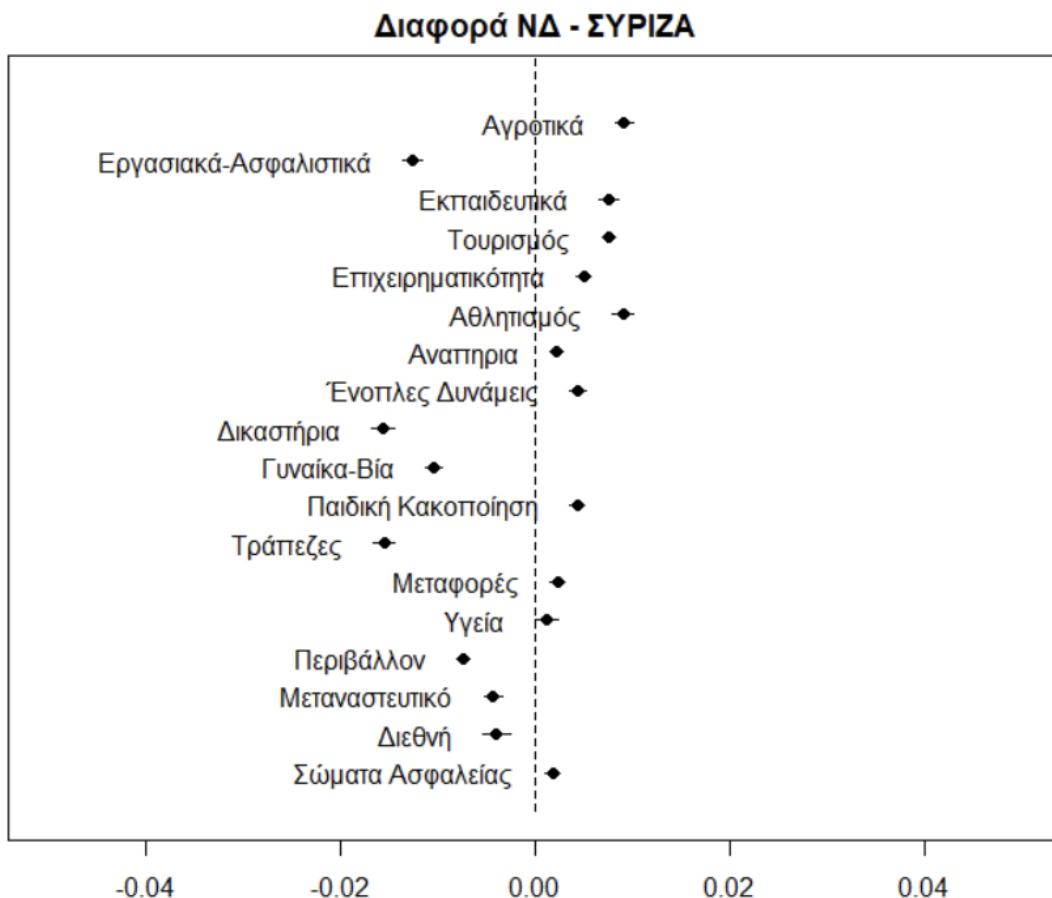


Figure 3 Comparison Between ND and SYRIZA per Topic Mentions

Figure 4 visualizes the comparative standing of New Democracy (ND) and PASOK across various political and social topics, using topic modelling. The points to the right indicate the issues in which New Democracy has involved more frequently than PASOK, while points to the left indicate topics where PASOK mostly focused on.

The figure reveals a landscape where New Democracy holds a broader advantage across most policy topics compared to its standing against PASOK. The data points to the right of the vertical zero-line show that ND is more favorably associated with sectors such as Agriculture, Tourism, Security Forces, and Education. PASOK maintains a lead in issues related to Justice, Women & Violence, and the banking sector. Furthermore, the difference between the two parties on the topics about Environment and Labor issues is insignificant.

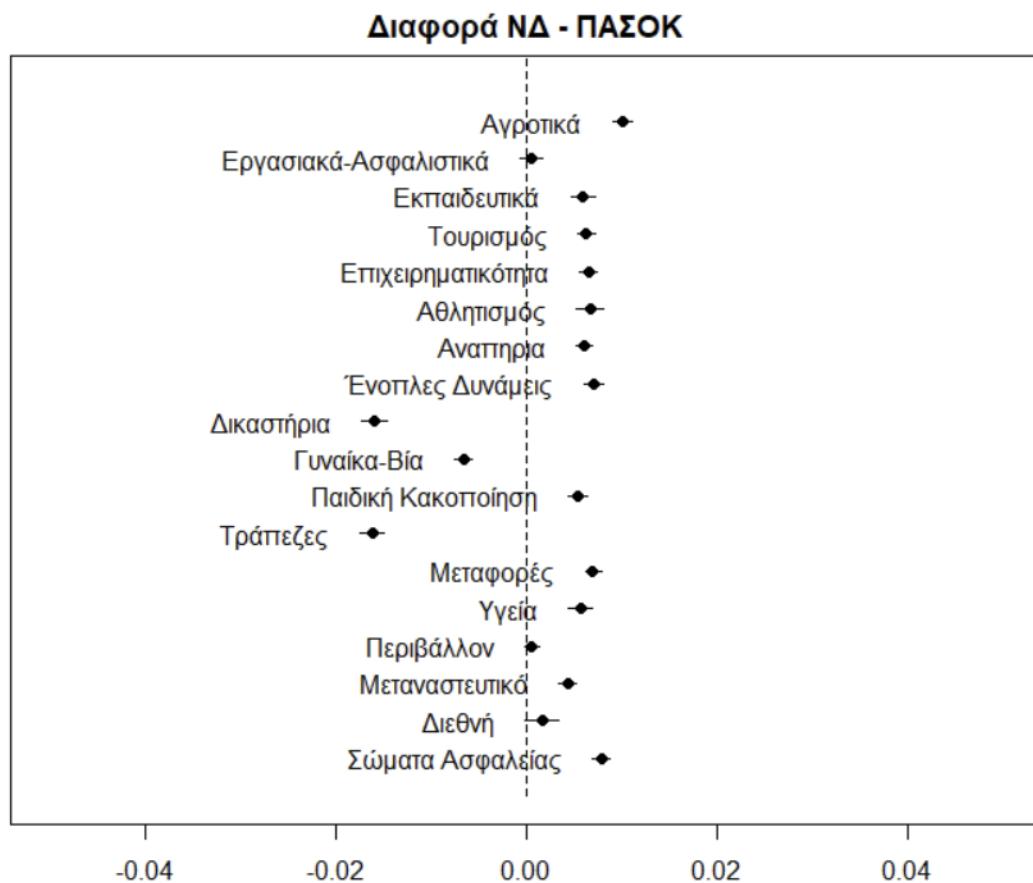


Figure 4 Comparison Between ND and PASOK per Topic Mentions

Figure 5 visualizes the comparison per topic mentioned between New Democracy (ND) and Greek Solution (Elliniki Lisi). More specifically, the figure visualizes the relative difference in thematic engagement between New Democracy and Greek Solution.

New Democracy focused in areas associated with administrative stability and economic development, Education, Tourism, Entrepreneurship, and Labor issues. In contrast, Greek Solution focused more on International Affairs, Justice and Migration, likely capitalizing on nationalist sentiments, as well as on the topic of Banks, suggesting an economic critique.

### Διαφορά ΝΔ - ΕΛΛΗΝΙΚΗ ΛΥΣΗ

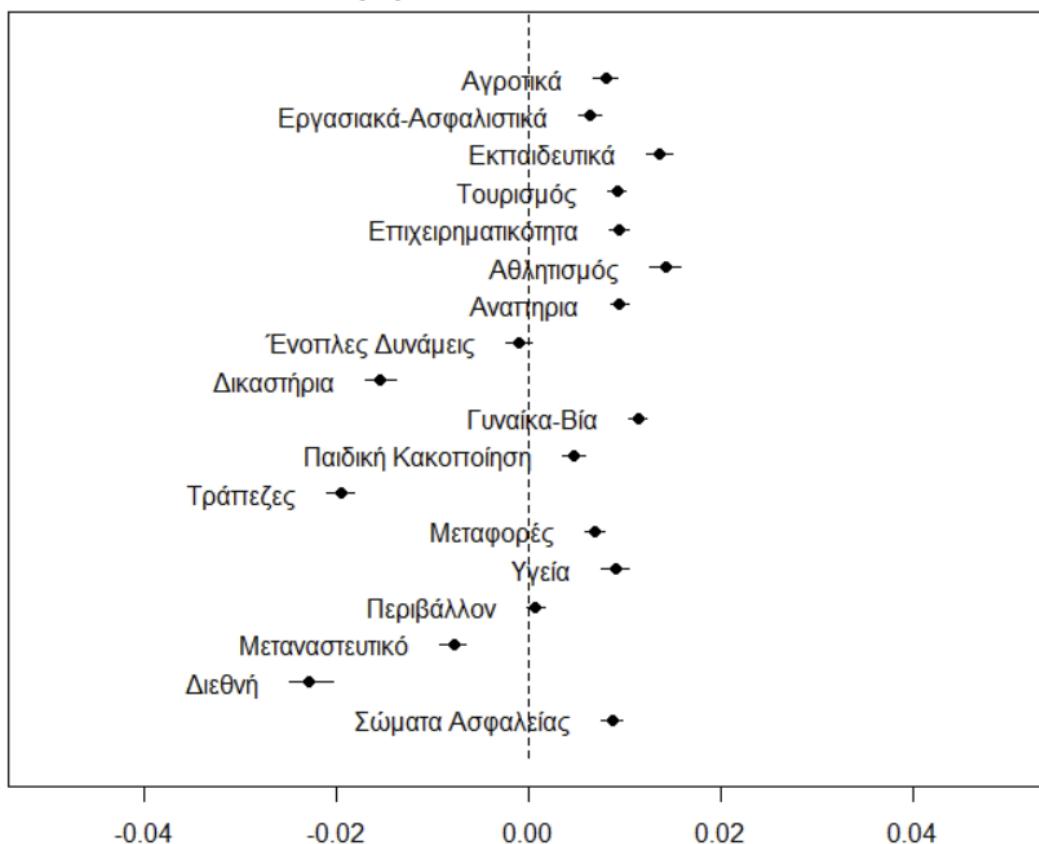


Figure 5 Comparison Between ND and Elliniki Lisi per Topic Mentions

Figure 6 reveals a highly distinct political profile for Pleusi Eleftherias compared to New Democracy. While New Democracy maintains its standard dominance on the majority of topics, it loses ground on specific issues tied to justice and advocacy.

The data figure reveals an overwhelming advantage for Pleusi Eleftherias in the category of Courts/Justice, an outlier so extreme it suggests that this party is seen almost exclusively through the lens of legal activism and institutional accountability. Even more notably, this is the only comparison where another party has focused more than New Democracy on Entrepreneurship, suggesting that Pleusi has successfully appealed to small business owners or professionals in a way other parties have not.

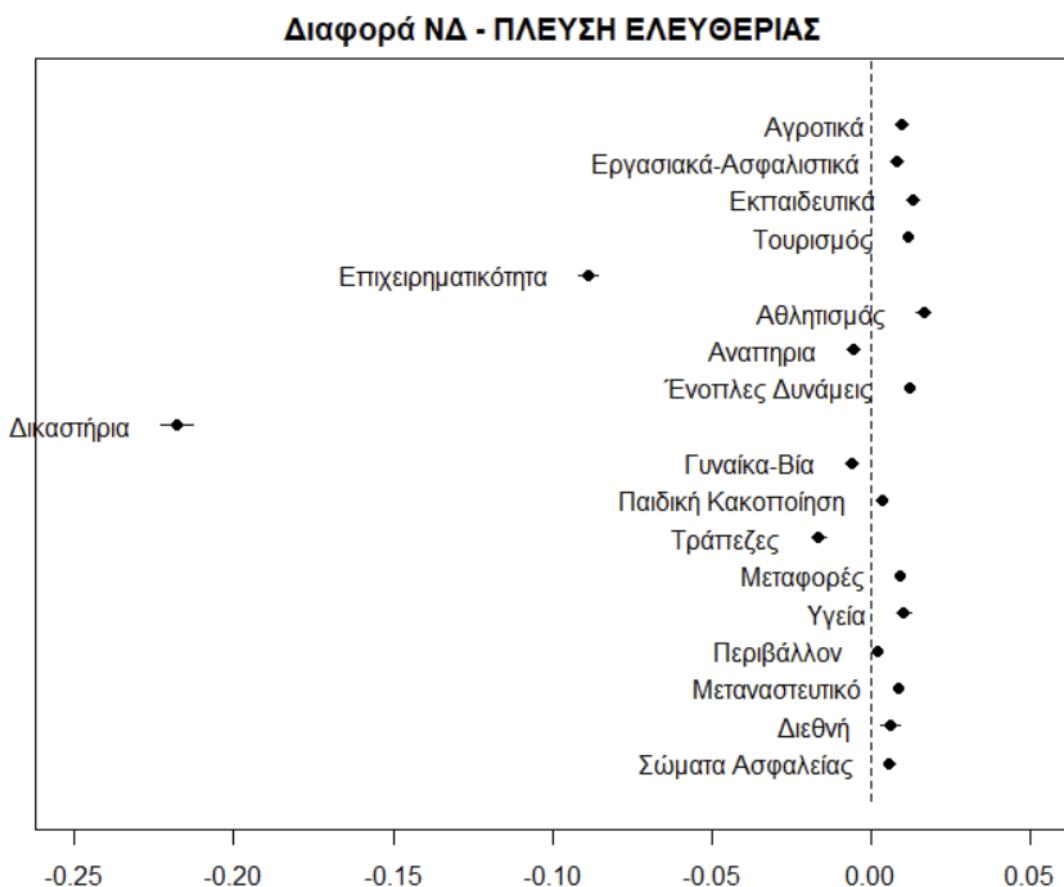


Figure 6 Comparison Between ND and Plefesi Eleftherias per Topic Mentions

Figure 7 visualizes the comparison between New Democracy (ND) and MeRA25 based on the topics mentioned. While New Democracy mostly focuses on governance-heavy sectors like Security, Tourism, Education, and Agriculture, MeRA25's opposition is characterized by a singular, intense focus on the financial sector. The data point for Banks indicates that MeRA25 essentially "owns" this topic in the context of this comparison, reflecting a party identity deeply rooted in critiques of the banking system and economic austerity. In addition, the figure reveals that MeRA25 more frequently mentioned about Environment, Justice, and social issues like Women's Rights and Labor, while conversely, ND focused on issues related to Agriculture, Education, Armed Forces and Tourism.

Διαφορά ΝΔ - ΜΕΡΑ25

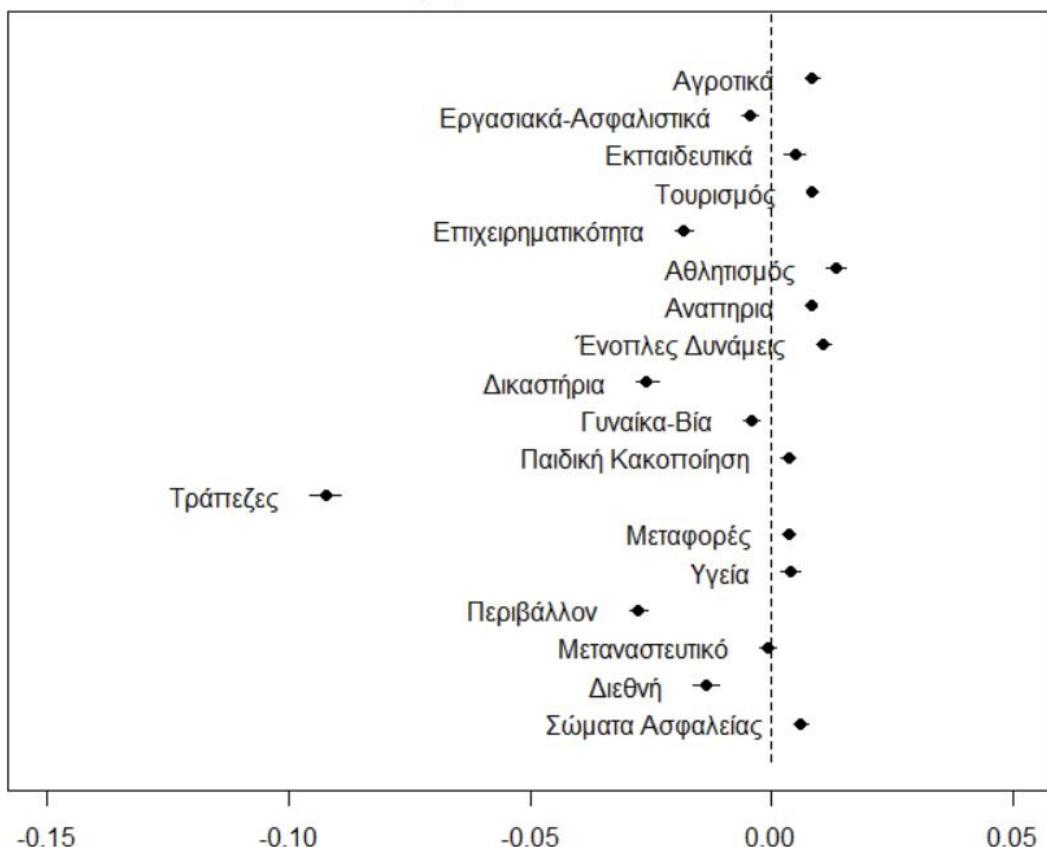


Figure 7 Comparison Between ND and MERA25 per Topic Mentions

## Sentiment Analysis

To evaluate the emotional valency of political discourse surrounding immigration, a normalized sentiment index was calculated for each Facebook post. In order to measure the sentiment score () we used two possible formulas.

### Positive-Negative/Positive+Negative+Neutral (Formula 1)

### Positive-Negative/Positive+Negative (Formula 2)

The sentiment analysis revealed that positive posts accounted for 54.4% of the total, followed by neutral posts at 28.5% and negative posts at 17.1%. We prefer to choose Formula 2, in order to emphasize the directional lean of the discourse by excluding neutral terms from the denominator. Had we utilized Formula 1, which includes neutral words in the denominator, the resulting percentage would have been lower due to the larger divisor. Consequently, the use of

Formula 2 provides a more focused measure of the net sentiment, highlighting the clear predominance of positive over negative framing in the analyzed content.

This metric yields a value ranging from -1 (purely negative) to +1 (purely positive), providing a comparative measure of the affective lean of each party's rhetoric.

## Immigration Topic Frequency by Political Party

The Figure 8 below presents the relative frequency of immigration-related references in Facebook posts published by major Greek political parties. By measuring how often immigration appears as a topic in each party's online communication, the chart provides insight into the salience of immigration within different partisan agendas and highlights variations in issue prioritization across the political spectrum.

Greek Solution (Elliniki Lisi) clearly stands out, exhibiting the highest relative frequency of immigration-related content. This indicates that immigration occupies a central position in the party's Facebook communication and functions as a core issue around which political messaging is structured.

MeRA25 ranks second in terms of immigration topic frequency. Although the issue is less dominant than in the case of Greek Solution, the relatively high frequency indicates that immigration remains an important component of MeRA25's online discourse.

SYRIZA occupies an intermediate position in distribution. Immigration-related references appear at a moderate frequency, implying that while the issue is clearly present in the party's communication, it does not dominate the agenda.

New Democracy (ND) and PASOK display similar and comparatively lower frequencies of immigration-related content. For both parties, immigration appears to play a secondary role in Facebook posts, indicating a more selective or contextual engagement with the topic.

Plefsi Eleftherias records the lowest relative frequency of immigration references among the parties included.

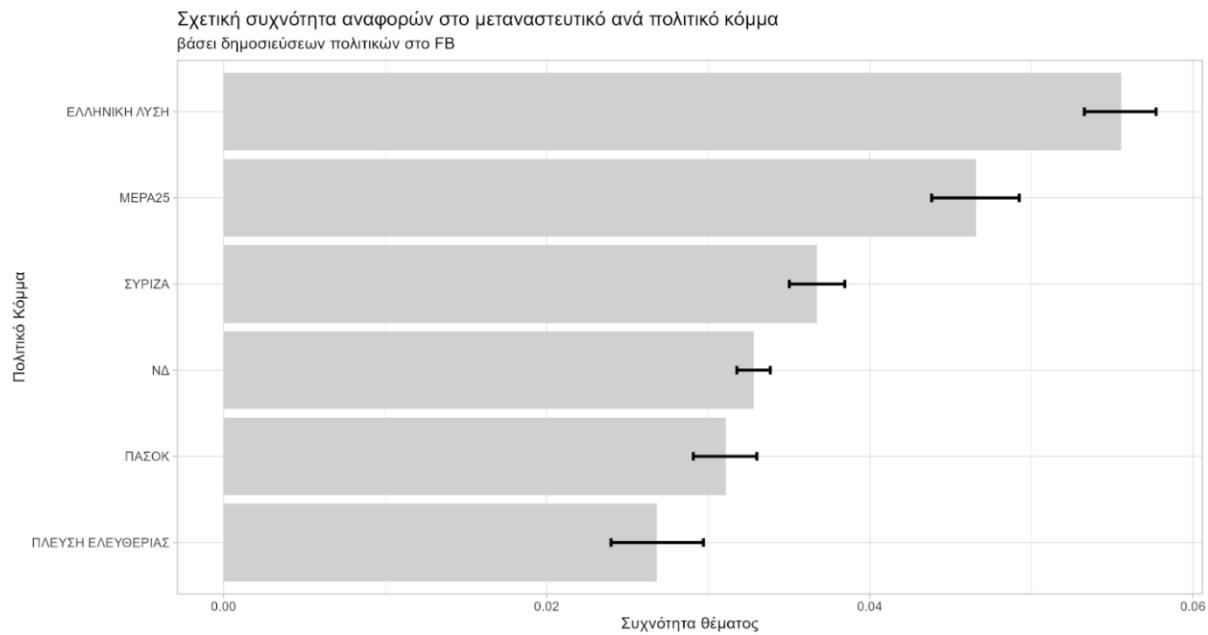


Figure 8 Frequency of Mentions of the Immigrant Topic by Political Party

## Comparative Sentiment Analysis: Greek Solution vs. SYRIZA and word-clouds

The analysis revealed a significant polarization in the framing of immigration issues following Formula 2. The maximum divergence in sentiment scores was observed between the right-wing populist party **Greek Solution (Elliniki Lysi)** and the left-wing **SYRIZA**. Specifically, Greek Solution exhibited a markedly negative mean sentiment score of **-0.296**, while SYRIZA maintained a positive mean score of **0.172**.



Figure 9. Word cloud of EL on immigration

These results align closely with the established ideological positions of the two parties regarding inclusive societies:

- **Greek Solution:** The negative score reflects a "securitization" framework. Qualitative analysis via word clouds indicates a high frequency of terms such as "illegal immigrants" and "borders." Their discourse consistently emphasizes a perceived lack of national security, frequently blaming the government for failing to protect the country's borders from what they frame as an external threat (Figure 9).
- **SYRIZA:** In contrast, the positive sentiment score reflects a "humanitarian" framework. Their rhetoric prioritizes inclusivity and human rights, with word clouds highlighting terms such as "human" and "hospitality". Notably, SYRIZA displays a linguistic preference for the term "refugees" rather than "immigrants," framing the issue as a matter of international protection and social responsibility rather than a security crisis.

This divergence demonstrates how lexical choices—such as the distinction between "illegal immigrant" and "refugee"—serve as indicators of broader ideological stances on social inclusion and the role of the state in managing migration.



Figure 10 Word cloud SYRIZA on immigration

# Concluding Remarks

The methodological approach adopted in this report provides a robust foundation for analyzing party positions across a wide range of policy domains, including those most relevant to the study of inclusive societies. By systematically coding every quasi-sentence of the available party manifestos according to the Euromanifesto framework, deliverable D3.3 contains the entirety of parties' programmatic statements at the occasion of the 2024 European Parliament election.

In addition, this report includes an analysis of social media posts. The analysis confirms that political parties systematically differentiate themselves through the issues they choose to emphasize, and that this differentiation is neither random nor merely communicative, but deeply strategic. Topic selection functions as a key mechanism through which parties signal identity, ideological orientation, and political priorities to the electorate. Especially for opposition parties, the choice of topics is closely tied to the objective of challenging the governing party by foregrounding policy areas that expose weaknesses, failures, or contradictions in government performance.

Across the dataset, opposition actors consistently concentrate on issues they perceive as electorally advantageous or morally authoritative, such as justice, labor and insurance, banking practices, migration, and institutional accountability. By contrast, the governing party (New Democracy) tends to emphasize areas associated with stability, administration, and economic development. This confirms that online political discourse reproduces classic patterns of issue ownership, while simultaneously intensifying them through the immediacy, reach, and low-cost dissemination afforded by social media platforms.

The findings from the sentiment analysis provide a quantitative confirmation of the profound ideological schism within the Greek party system regarding the transition toward an inclusive society. By examining the affective lean and lexical choices of political actors, we can distinguish between the "friends" and "foes" of inclusive discourse based on their framing of marginalized groups. The radical right, exemplified by Greek Solution, functions as a vocal antagonist to inclusivity; by utilizing a securitization lens and negative sentiment, they construct a "foe" narrative that portrays migration as a threat to national integrity. Their rhetoric focuses on exclusion, emphasizing the enforcement of borders and the illegality of the "other."

Conversely, the radical left, represented by SYRIZA, adopts the role of an advocate for inclusive societies, framing immigration through a humanitarian and rights-based lens. Their preference for the term "refugee" over "immigrant" and the high frequency of the word "humans" signals a deliberate attempt to foster empathy and social solidarity. This ideological divergence suggests that while the radical right views inclusivity as a zero-sum game that threatens the native population, the radical left presents it as a normative necessity for a democratic state.

Ultimately, this study demonstrates that on the issue of immigration, the "sentiment gap" between the radical left and right is not merely a difference in tone, but a fundamental disagreement on the boundaries of the political community. These polarized communication strategies serve to mobilize distinct electorates: one through the politics of fear and exclusion, and the other through the politics of compassion and universal human rights.

Overall, this report provides researchers, policymakers, and stakeholders with a comprehensive and methodologically sound tool for examining how political parties articulate, prioritize, and frame issues linked to social inclusion (among others). In so doing, it contributes to a deeper understanding of the political dynamics that underpin inclusive societies.

## References

Braun, D., & Reinl, A.-K. (2022). *European Parliament Elections – Euromanifesto Study (Documents 2019)* (No. ZA7892; Version 1.0.0) [Dataset]. GESIS. <https://doi.org/10.4232/1.13977>

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>

Kang, B., Song, M., & Jho, W. (2013). A Study on Opinion Mining of Newspaper Texts based on Topic Modeling. *Journal of the Korean Society for Library and Information Science*, 47(4), 315–334. <https://doi.org/10.4275/KSLIS.2013.47.4.315>

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.

<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, 111(515), 988–1003. <https://doi.org/10.1080/01621459.2016.1141684>

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91, 1–40.

<https://doi.org/10.18637/jss.v091.i02>

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064–1082.

<https://doi.org/10.1111/ajps.12103>

Robinson, & Silge, J. (2017). *Welcome to Text Mining with R | Text Mining with R*. O'Reilly Media, Inc. <https://www.tidytextmining.com/>

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*.

Volkens, A., Bara, J., Budge, I., & McDonald, M. D. (Eds.). (2014). *Mapping Policy Preferences from Texts: Statistical Solutions for Manifesto Analysts*. Oxford University Press.