Sentiment and Thematic Analyses of British Politician Speeches

Kanav Bhagat - 730013149
Department of Computer Science
University of Exeter, Exeter, UK

***Abstract*—This project examines the language used in speeches by British politicians, aiming to uncover the emotions and topics they frequently discuss. By analysing a collection of speeches from UKPOL.co.uk [1], we identify prevailing sentiments and themes across different political parties and compare these findings with public opinions from YouGov.co.uk [2] to see how well they align.**

I. Introduction

Politicians give speeches to share their messages, influence what people think, and explain their plans. Our project looks into the last 100 speeches to find out what feelings they express and what subjects they talk about. The subjects can vary from issues like Defence, Education or Economy. By studying the language politicians use, we can learn more about what they are trying to say and how their words connect with the public.

A. Motivation

As political landscapes become increasingly polarised, understanding the language of politics is crucial. This project is driven by the potential of sentiment and thematic analysis to dissect political speeches, offering clarity on politicians' emotional appeals and focal issues. Further, it also compares the results of the experiments with the trackers from *YouGov* to see if these speeches actually affect the public opinions or not.

II. Related Work

My project builds upon significant previous research in the area of sentiment analysis within political contexts, drawing inspiration from two pivotal studies:

- Karpouzis et al.'s [3] exploration of social media communication by U.S. politicians demonstrated the power of sentiment analysis in distinguishing political communication styles. Their approach, using spaCy for sentiment detection, laid the foundation for our sentiment analysis framework. Inspired by their methodology, my project employs the VADER (Valence Aware Dictionary and Entiment Reasoner) sentiment analysis tool. VADER is particularly suited for social media text but proves equally effective in analysing the sentiment of political speeches, providing a lexicon and rule-based framework designed for the sentiment analysis.
- Rodríguez-Ibáñez et al.'s [4] work on sentiment analysis during the Spanish elections illuminated the shifting sentiments within political discourse over time. Their use of advanced statistical techniques to analyse tweet sentiments across an electoral period highlighted the dynamic nature of public opinion. This insight was crucial in my thematic exploration, where we adopted Latent Dirichlet Allocation (LDA) for thematic analysis. LDA, a machine learning algorithm, enables us to identify underlying themes across a large corpus of political speeches, offering a systematic approach to uncovering the thematic structure of political discourse.

These studies were instrumental in shaping the direction and methodology of my project. The use of VADER allowed us to conduct a detailed sentiment analysis, assessing the emotional tone of speeches across British political parties. Concurrently, the application of LDA for thematic analysis provided a robust mechanism to categorise speeches into distinct themes, revealing the core subjects addressed by politicians.

III. Dataset

For the project, I created a detailed dataset of the 100 most recent speeches from UKPOL.co.uk. This dataset was essential for understanding the various aspects of political speech in the UK. To gather this data efficiently, I developed a Python script named *GetSpeeches.py* that utilised the BeautifulSoup library, a powerful tool for web scraping and parsing HTML.

The Python script worked by visiting *www.ukpol.co.uk* [1] to access a list of the latest speeches. Through BeautifulSoup, the script was able to read through the website's HTML and identify the important parts of each speech:

- **Politician Name** and **Speech Title**: Found within the webpage's *h1* tags. I split these elements at the dash ('–') to separately capture the politician's name and the speech title.
- **Speech URL**: Extracted from the link element (*a* tag), allowing me to locate the full text of each speech.
- **Speech Text**: Pulled from the *div* tag identified by the class *cm-entry-summary*, which contained the entire text of the speech.

After collecting this information for each speech, I organised it into a structured format, saving it into a CSV file. This file includes columns for 'Politician Name', 'Speech Title', 'Speech URL', 'Speech Text', and 'Party' — the last of which I added based on each politician's known party

affiliation to ensure a comprehensive analysis across political parties.

Refer Table I to check structure of csv file.

TABLE I
STRUCTURE OF DATASET CSV

Politician Name	Speech Title	Speech URL	Speech Text	Party

The result, stored in *SpeechData_with_Party.csv*, provides a rich source of data for the analysis. This dataset serves as the foundation for our exploration into the sentiment and thematic content of political speeches, offering insights into how different parties and politicians articulate their messages.Please note: The dataset contains **85** Speeches from Conservative Party, **7** speeches from Labour party and **8** speeches from Others. This is highly skewed since the conservative party is in power as of conducting experiments and makes more speeches in British Parliament.

IV. METHODOLOGY

The methodology was split into two main parts: **Sentiment analysis** and **Thematic analysis**.

A. Sentiment Analysis

I chose the ***VADER*** sentiment analysis library for evaluating the emotional tone of political speeches. VADER is specially designed for texts from social media, which often resemble the style of political speeches in their expressiveness. Moreover VADER doesn't require any pre-processing as it has the ability to analyse raw text directly [5]. Moreover VADER gives us four scores: positive, neutral, negative, and a composite score. The composite score indicates the overall sentiment, ranging from -1 (very negative) to +1 (very positive).

B. Thematic Analysis with LDA

Next, I aimed to uncover the key topics within the speeches through thematic analysis, using the Latent Dirichlet Allocation (LDA) [6] model. This part involved several steps:

- Pre-Processing: Refer Fig 1 for Pre-processing pipeline.

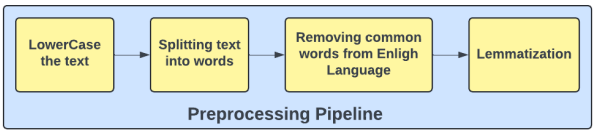

Fig. 1. Diagram Showing Pre-processing pipeline

- Building the Corpus and Dictionaries: Post-preprocessing, I created a corpus (a collection of the processed texts) and dictionaries (which map each word to a unique ID) to structure the data for the LDA model.
- Applying LDA: With the data prepared, we used the LDA model to identify groups of words that represent themes. This step revealed the main subjects that politicians talk about in their speeches.

V. RESULTS AND DISCUSSION

The comprehensive analysis of the political speeches dataset has led to distinct insights into the sentiments expressed by politicians and the prevalent themes within their speeches. The findings are categorised into sentiment analysis and thematic analysis, as detailed below.

A. Results for Sentiment Analysis

The sentiment analysis, conducted using the VADER library, revealed a diverse range of emotional tones across the political speeches. VADER provided us with four key metrics for each speech: positive, neutral, negative, and a composite score, which together offer a complete picture of the sentiment landscape.

- **Overall Sentiment**: The analysis showed a predominance of positive sentiment across the speeches, with a composite average score of **0.808**, indicating a generally positive. This can be due to the upcoming elections and all the parties want to show a positive sentiments.

- **Party-wise Sentiment Breakdown**: We observed notable differences in sentiment scores among political parties. For instance, the Conservative Party's speeches had an average composite score of **0.884**, while the Labour Party's speeches showed an average composite score of **0.467**. This variance highlights the differing emotional appeals used by the parties in their communication.

  The Conservative's high positivity likely reflects a strategic focus on optimism and achievements, aiming to portray confidence and a positive vision. Labour's more moderate score suggests a balanced rhetoric that acknowledges both successes and challenges, aiming to resonate with a broader audience by blending hope with realism. The lowest scores among Other parties may indicate a critical stance towards the status quo, focusing on unaddressed issues and advocating for change.

- **Most Positive/Negative Speech**: The speech titled *"2024 Budget Speech"* by Jeremy Hunt from the Conservative Party stood out with the highest positive composite score of **1.00**, whereas *"2024 Speech on Freedom and Democracy in Iran"* by Bob Blackman recorded the lowest composite score of **-0.998**.

  Jeremy Hunt's "2024 Budget Speech" scored a perfect 1.00, showing optimism about future economic

plans under the Conservative Party. In contrast, Bob Blackman's speech on Iran scored -0.998, reflecting serious concerns about freedom and democracy in a critical tone.

B. Results for Sentiment Analysis

The thematic analysis, utilising the Latent Dirichlet Allocation (LDA) model, identified key themes within political speeches from the Conservative and Labour parties, classified into distinct topics:

Conservative Party Themes:

- Employment: Highlighted by discussions on "business" and "people," indicating a focus on job creation and economic growth.
- Defence and Security: Evident from mentions of "defence" and "northern_ireland," pointing towards national security and regional stability concerns.
- Law and Order: Themes related to "gambling" and "consultation" reflect on regulatory efforts and governance.

Labour Party Themes:

- Women: The frequent mention of "woman" and "service" suggests a concentration on women's rights and services.
- Education: Although not directly mentioned, themes around "government" and "funding" may imply a focus on educational funding and policy.
- Law and Order: Discussion on "authority" and "funding" hints at governance and social care, aligning with law and order through a social welfare lens.

VI. Public Opinion Vs Results

A. Economy

Based upon public opinion data acquired from textitYouGov.co.uk [2], **Labour party** has a slight lead over **Conservative party** as Showsn in Fig2

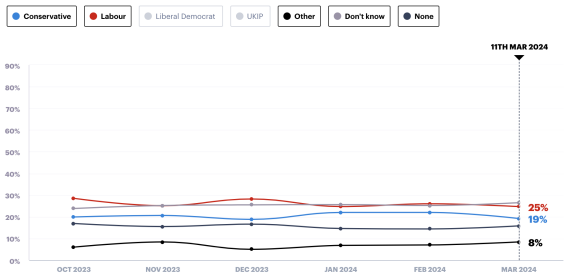

Fig. 2. Which political party would be the best at handling Economy?

B. Security

In terms of defence and security, public thinks **Conservative party** will do a better job compared to **Labour party** as shown in Fig 3

C. Unemployment

Public thinks **Labour party** will be better in terms of providing employment to people compared to **Conservative party** as shown in Fig 5

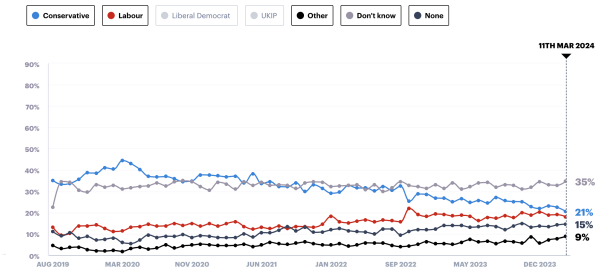

Fig. 3. Which political party would be the best at handling Defence and Security?

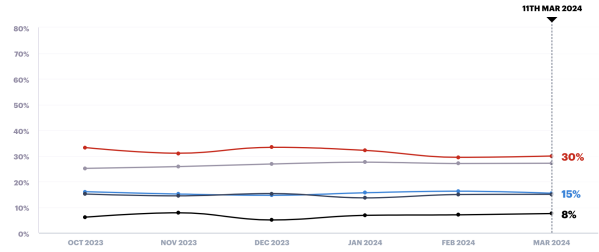

Fig. 4. Which political party would be the best at handling Unemployement?

D. Law and Order

In terms of Law and Order, public thinks **Labour party** will do a better job compared to **Conservative party** as shown in Fig 3

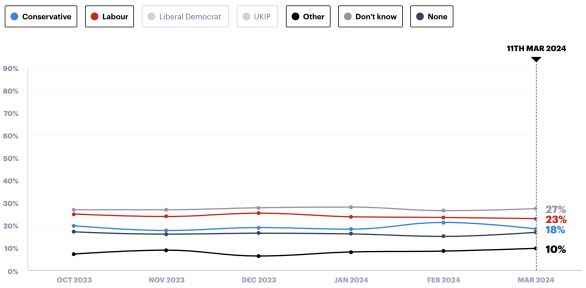

Fig. 5. Which political party would be the best at handling Unemployement?

VII. Conclusion

My study of British political speeches found that all parties, especially the ruling party, keep their language hopeful and positive. However, this approach doesn't seem to change what people think about the government's handling of the economy. Despite trying to sound optimistic, the government's struggles with economic issues are still a big concern for the public. This shows that just being positive in speeches is not enough to improve people's views on the government's performance, especially regarding the economy.

REFERENCES

[1] "Ukpol.co.uk," https://www.ukpol.co.uk/.

[2] "Yougov.co.uk," https://yougov.co.uk/.

[3] K. Karpouzis, S. Kaperonis, and Y. Skarpelos, "Identification of common trends in political speech in social media using sentiment analysis," 10 2022.

[4] X. Zuo, B. Zhao, H. Li, J. Liu, B. Lu, and X. Liu, "Influence of key parameters on the performance of the improved repetitive ipps with iccos," *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 8, pp. 1–5, 2021.

[5] R. Malde, "A short introduction to vader," June 2020, [Online; posted 8-June-2020]. [Online]. Available: https://towardsdatascience.com/an-short-introduction-to-vader-3f3860208d53#:~:text=Preprocessing.&text=Unlike%20with%20some%20supervised%20methods,it%20will%20determine%20the%20sentiment.

[6] M. F. P. Gillies, D. Murthy, H. Brenton, and R. Olaniyan, "Theme and topic: How qualitative research and topic modeling can be brought together," *ArXiv*, vol. abs/2210.00707, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:252683229