
Analyzing Turkish Parliamentary Speeches: Insights from the TBMM Corpus

M. Burak Kizil^{* 1}

1. Introduction

Parliamentary debates often reflect the core political, social, and economic issues of a country. They offer a window into how elected representatives articulate policy priorities, respond to crises, and negotiate competing ideologies. In Turkey, the TBMM (Türkiye Büyük Millet Meclisi) is a prime arena where these discussions unfold. This study proposes to investigate a comprehensive corpus of Turkish parliamentary texts made available through a public GitHub repository. By exploring topics, sentiments, and rhetorical patterns in TBMM debates, we aim to uncover trends, shifts, and underlying structures in legislative discourse.

2. Research Questions

We intend to address a range of questions that revolve around content, evolution, and sentiment in the Turkish parliamentary record:

Topic Evolution

- Which policy areas (e.g., economy, education, foreign affairs) receive the most attention, and how has their prominence changed over time?
- Do certain political parties or individual deputies consistently focus on specific themes?

Sentiment and Rhetorical Style

- How do sentiment (positive, neutral, negative) and emotional tone vary across different parties or specific time periods?
- Are there discernible patterns of polarizing or conciliatory language around major legislative proposals or national events?

Temporal and Event-Driven Shifts

- How do significant events (e.g., economic crises, constitutional referendums) influence the frequency of certain keywords, topics, or tones in parliamentary speeches?
- Can we identify structural breaks or turning points in discourse that coincide with political transitions or legislative changes?

3. Data

Corpus: The “turkish-parliament-texts” repository ([GitHub Link](#)) provides access to transcripts from the TBMM. According to its documentation, it includes speeches from various legislative terms, capturing debates on key issues and legislative proceedings.

4. Methodology

Our proposed methodology blends both traditional text analytics and advanced machine learning to gain robust insights. While we outline core steps below, we may adapt or expand them as new information about the data emerges.

Data Preparation: We plan to remove extraneous formatting, handle punctuation and special characters, and tokenize the text using Turkish language libraries (e.g., Zemberek). Lemmatization or stemming will be considered to reduce words to their base forms, especially important in an agglutinative language like Turkish.

Topic Modeling & Thematic Analysis:

- **Unsupervised Learning:** We intend to implement topic modeling methods (e.g., LDA – Latent Dirichlet Allocation, or neural-based approaches like BERTopic) to identify key themes.
- **Party or Speaker-Level Insights:** We will examine the distribution of topics across parties, over legislative years, and around major events (as identified by session dates).

Sentiment Analysis: We will experiment with Turkish sentiment lexicons or train supervised models (e.g., logistic regression, Transformer-based classifiers) on annotated subsets if we can source or create labeled data.

Refinement & Advanced Techniques: As the data structure becomes clearer, we will decide whether supervised classification (e.g., training a party affiliation classifier based on speech text alone) or clustering (to identify emergent discourse coalitions) offers additional value. We may incorporate named entity recognition (NER) to spot references to key people, places, or institutions, potentially shedding light on how different parties position themselves on critical issues.

5. Expected Contributions

This research will enhance understanding of Turkish parliamentary discourse by demonstrating how advanced text analytics and machine learning methods can uncover hidden patterns, themes, and sentiment shifts. Substantively, it may reveal how deputies align or diverge across topics, the rhetorical tone surrounding contentious legislation, and the impact of significant national events on parliamentary debates. Methodologically, the study will serve as a blueprint for analyzing large-scale textual data in other legislative or policy contexts, especially in low-resource languages like Turkish, where specialized NLP tools are rapidly evolving.