

A Comparative Text Analysis of Immigration Coverage in Liberal vs. Conservative Media

Yuqi Chen, Haoyuan Deng, Kexin Huang, Lin Pan; John Rachlin

Background

Public debates on immigration policy are often polarized by ideology, and media arguments and rhetoric are shaped accordingly. In this project, we built an extensible framework in Python (GovSnatch) for comparing and visualizing the content of more than 10 immigration-related articles from liberal and conservative news sources. We explored differences in diction, double entendre, topic emphasis, and sentiment in an attempt to highlight how political leanings influence media narratives

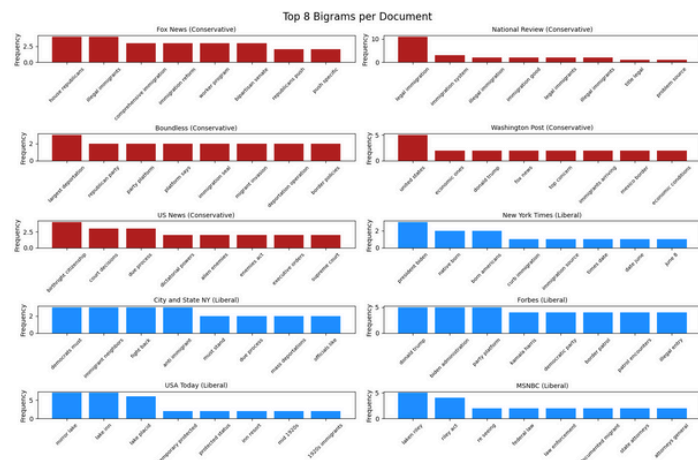


Figure 1: Bar Plot of Articles's Top 8 Bigrams

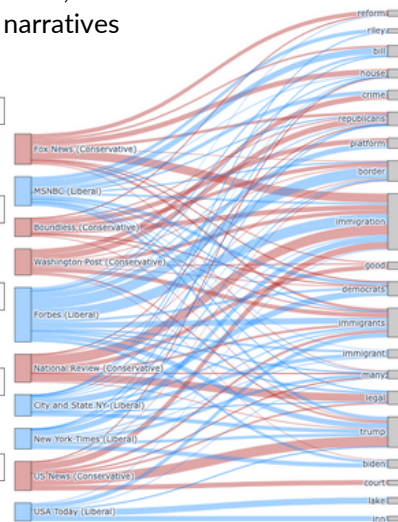


Figure 2: Sankey Diagram, Mapping between Articles and 3 Most-frequent Words

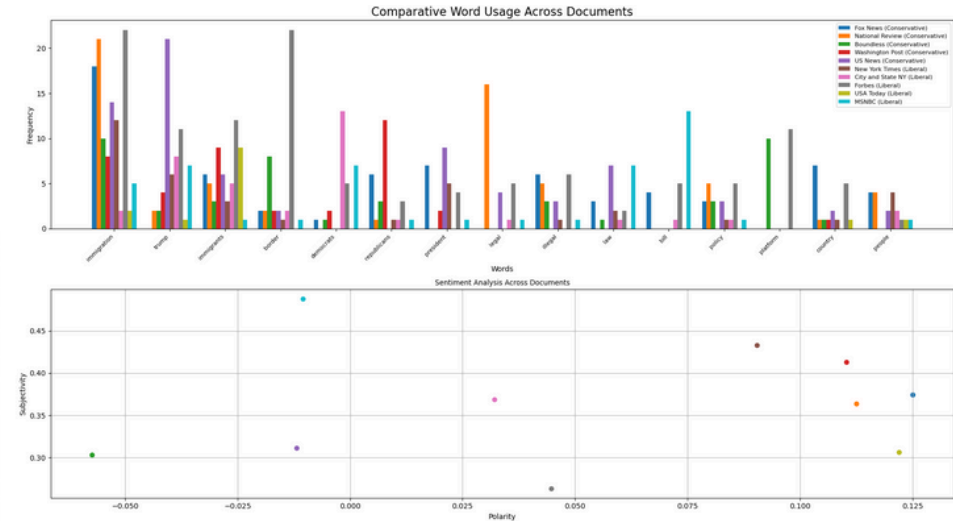


Figure 3: Bar chart of top 15 words

Process and Methods

1. Parsing & Cleaning

- Converts text to lowercase, removes punctuation, splits into tokens (words), and filters out common stopwords using stopwords.txt.

2. Feature Extraction & Metrics

- We store, for each article, a results dictionary with:
 - wordcount: frequency of each word
 - numwords: total count of words (after stopwords removal)
 - bigramcount / trigramcount: frequency of each pair/triple of consecutive words
 - avg polarity / avg subjectivity: sentiment metrics computed via TextBlob
 - avg sentence length: mean number of words per sentence
 - readability score: average word length

3. Visualizations

- Sankey Diagram: Links each article to its most frequent words (Figure 1).
- Bigram Subplots: Shows the top bigrams per article (Figure 2).
- Comparative Chart: Bar chart for frequency of top words (Figure 3) and scatterplot for sentiment (Figure 4).

Conclusion and Next Steps

The framework quantifies the stark rhetorical differences between conservative and liberal immigration discourse, revealing fundamentally different linguistic frameworks that reflect divergent policy priorities and values. These patterns suggest media outlets actively shape immigration narratives through strategic word choice and framing rather than merely reporting facts. To next steps, we will expand our dataset to include international sources and social media, enhance GovSnatch with advanced semantic analysis and topic modeling capabilities, and conduct longitudinal analysis tracking rhetorical shifts around major policy events.

Figure 4: Scatter plot of polarity vs. subjectivity