

Semantic Polarization in Broadcast News (2015–2025)

Shi Qiu
Virginia Tech
Alexandria, VA, USA
shiqiu@vt.edu

Zechen Li
Virginia Tech
Alexandria, VA, USA
lzechen@vt.edu

1 Project Idea

We will investigate semantic polarization in broadcast news language, inspired by the first 2 studies conducted by Ding *et al.*[1]. Our research aims to reproduce and expand upon the original study by analyzing updated data (2015-2025) and applying the latest NLP methods. We will quantitatively compare whether the conclusions drawn from the original study using data up to 2020 will still be valid in 2025, and whether the new NLP methods can yield more accurate conclusions. Specifically, we will conduct:

Study 1: This study focuses on quantifying the growing partisan media gap by measuring how CNN and Fox News discuss identical keywords over a decade. The core implementation involves fine-tuning a RoBERTa model on an sentiment dataset[3]. This classified model is applied to the corpus to generate a time-series of affective bias, which allows us to track the evolution of media framing bias using NLP.

Study 2: Identify which words(tokens) characterize the differences in how CNN vs. Fox discuss those keywords. We will train classifiers to predict if a news piece came from CNN for Fox, and then interpret the model to find the tokens that mostly contributes to each classifiers. By comparing different classifiers, we can test the robustness of these linguistic markers.

2 Motivation

Original research measured polarization using BERT embeddings, but failed to explicitly quantify the emotional or affective stance of the two networks. Our primary motivation is to fill this gap by introducing the Supervised Sentiment Polarization metric. SSP provides a concrete, interpretable value that directly reflects the media’s affective bias[2]. This is highly relevant as affective framing significantly influences audience perception.

Additionally, we want to examine whether using different embedding models and classifiers can improve the detection of semantic differences or provide any new insights. We hope to examine whether the results in the original paper depend on a specific method by replacing the model used in the original paper, and whether replacing the model can bring better performance or interpretability.

Furthermore, the original study’s data ended in 2020. During the Q&A session of our paper presentation, questions arose regarding whether the results would change with the data scope. This prompted us to update the analysis to 2015-2025: Since 2020, has this polarization trend continued, intensified, or reversed? By extending the time frame, we can test the robustness of the original study’s results and

observe whether new events (such as election results and policy changes) affect semantic polarization.

3 Dataset

Our dataset consists of TV news transcripts from CNN[4] and Fox News[5] covering the years 2003 through 2025. We will only use data after 2015 to limit the project scope. Ultimately, we filtered out approximately 145,000 CNN transcripts and 40,000 Fox News transcripts.

CNN transcripts are stored in structured CSV format. Each row contains a headline, the full text, and some metadata fields including year of publishing.

Fox News transcripts are

Following the original study, we will focus on transcripts that mention any of several politically charged keywords across three core topics:

- *Racism*: e.g., occurrences of “racism” or “racist”
- *Law Enforcement*: terms like “police”
- *Immigration*: terms “immigration” or “immigrant”

4 Example for Dataset

Table 1: Sample Fox News Transcripts (2022)

Date	Guests
2022-05-29	Mo Brooks, Olivia Beavers, Ben Cardin
2022-05-22	Brian Deese, Julie Pace, Ronna McDaniel
2022-04-17	Kevin McCarthy, Ashish Jha

Table 2: Sample CNN Transcripts (2022)

Date	Topic Coverage
2022-05-29	Immigration policy debate
2022-05-22	Law enforcement relations
2022-04-17	Racism and social justice

Table 3: Dataset Statistics (2015-2025)

Metric	CNN	Fox News
Total Transcripts	~145,000	~40,000
Coverage Period	2003-2025	2003-2025
Analysis Focus	2015-2025	2015-2025

Table 4: Target Keywords by Topic Category

Topic	Keywords
Racism	“racism”, “racist”
Law Enforcement	“police”
Immigration	“immigration”, “immigrant”

Table 5: Keyword Usage Examples

Network	Example Context
<i>Immigration:</i>	
Fox News	debate over policy
CNN	immigrant families
<i>Police:</i>	
Fox News	recruitment challenges
CNN	reform accountability
<i>Racism:</i>	
Fox News	political accusations
CNN	systemic issues

Table 6: Data Fields Used in Analysis

Field	Description
Title	Episode title
Publication Date	Date published (YYYY-MM-DD)
Description	Episode summary and guests
Full Text	Complete transcript text
Year	Publication year
Network	CNN or Fox News

5 Studies

Our project will be divided into two parts corresponding to Study 1 (semantic polarization over time) and Study 2 (finding key words to distinguish CNN and Fox News) from the original paper, with improved or alternative NLP methods for each.

5.1 Study 1

This is the core innovative contribution, directly measuring polarization as the absolute difference in the affective (sentiment) stance adopted by the two networks towards the keyword.

Sentiment Classification: A state-of-the-art **RoBERTa**-base model will be fine-tuned on an external, sentiment dataset to classify each sentence as having a Positive or Negative affective score[3].

Polarization Metric (SSP): We calculate the average sentiment score (\bar{S}) for each network and define the SSP as the absolute difference between these averages:

$$SSP_{w,t} = |\bar{S}_{CNN}w,t - \bar{S}_{Fox}w,t| \quad (1)$$

5.2 Study 2

In Study 2, we will reproduce the identification of which tokens distinguish between CNN and Fox reports when discussing the same topic. In the original study, the authors trained BERT-based classifiers (one per topic) using 2020 data to predict whether a given sentence came from CNN or Fox. They then used integrated gradient (IG) to find the tokens that contributed most to each classification. To this end, we plan to try several different classification methods and compare their results.

When performing the classification task, we first reproduce the BERT classifier implemented in the original paper. By comparing the 2020 data with the metrics from the original study, we can verify whether we have accurately reproduced the implementation in the original paper. As a comparative method, we additionally train a simpler SVM classifier using TF-IDF on the same task. While the linear model is not as powerful as BERT, it is more interpretable and can directly indicate which words have a stronger effect on the classification task. This can serve as a baseline for plausibility and interpretability. We expect that even the logistic model will capture a clear signal. Finally, we plan to experiment with improved BERT-based models, such as RoBERTa[3], to test whether they can improve performance on classification tasks.

Then, we apply IG to the best-performing classifier, which assigns an importance score to each word(token) in the input sentence, indicating the token’s “CNN” bias in the model’s predictions. By averaging these attribution results across multiple samples, we can determine which tokens in each topic have the strongest linguistic association with CNN and which have the strongest linguistic association with Fox. We will reproduce the results from 2020 to verify the accuracy of our obtained classification tokens. Then, we will extend our research to other years from 2021-2025, performing attribution on data for each year to further examine whether the distinguishing tokens change over time.

6 Metrics

In this section, we summarize the metrics we use to determine the performance of each study.

6.1 Study 1

Sentiment Polarization Time Series: Generate the annual SSP score for each keyword across the decade.

Coherence & Correlation: We will compute the SSP scores against the previously established Semantic Polarization scores to analyze the internal correlation between these two distinct metrics.

Success Measurement: Study 1 is successful if the SSP time series demonstrates a clear, increasing polarization trend, corroborating the SP baseline and known national political events.

6.2 Study 2

In Study 2, we will use standard classification metrics to ensure that our models can distinguish between CNN and Fox statements. We will reserve a subset of data for each topic as a test set. The metrics we will use include accuracy and F1-score. The BERT classifier in the original study achieved approximately 70-75% accuracy, indicating that this is a learnable but not easy task. If our best model achieves similar or higher performance (e.g., $\geq 75\%$ accuracy) on each topic subset, we will consider our reproduction successful. As a baseline comparison, we expect TF-IDF performance to be slightly lower, but if the two are similar, it is worth noting.

Since a key outcome of Study 2 is the list of distinguishing tokens for each network, we will evaluate the meaningfulness of these tokens. We will examine whether the top tokens identified by our method align with intuition or historical records. For example, if our 2020 analysis (using BERT and IG) produces the same types of tokens reported in the original paper, such as describing Fox as "illegal" (referring to immigration)

or CNN as "family," then we consider this validating of our method. We can roughly measure the overlap between the top 10 words in our model and the top 10 words in the paper (if any). The ultimate criterion for success is whether our findings can explain the polarization calculated in Study 1.

References

- [1] Xiaohan Ding, Michael Horning, and Eugenia H. Rho. 2023. Same Words, Different Meanings: Semantic Polarization in Broadcast Media Language Forecasts Polarity in Online Public Discourse. *Proceedings of the International AAAI Conference on Web and Social Media* 17, 1 (Jun. 2023), 161–172. doi:10.1609/icwsm.v17i1.22135
- [2] B. Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* (2012).
- [3] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL] <https://arxiv.org/abs/1907.11692>
- [4] Gaurav Sood. 2017. CNN Transcripts 2000–2025. doi:10.7910/DVN/ISDPJU
- [5] Gaurav Sood. 2022. Fox News Transcripts (2003–2025). doi:10.7910/DVN/Q2KIES