# Heard or Halted?
# Analyzing the Emotional Tone of Judicial Interruptions
## PPOL6801 Text As Data - Final Project

TIAN TONG

## 1. Abstract

This project investigates the role of gender in Supreme Court oral arguments by analyzing how interruptions affect the semantic content and emotional tone of advocates' speech. Using the ConvoKit Supreme Court Corpus (2010–2019), we extract 12,663 speech chunks from advocate-justice interactions. We address two central questions: (1) Do interruptions shift the semantic meaning of an advocate's argument? (2) Do interruptions directed at female advocates contain more negative sentiment than at male advocates? We employed GloVe-based embeddings to measure semantic shifts and apply lexicon-based sentiment analysis for emotional valence. We found that while semantic similarity remains high between speech before and after interruptions, and interruptions toward female advocates tend to have higher negativity. These findings contribute to ongoing discourse on gendered dynamics in elite deliberative settings and demonstrate the value of computational linguistics in empirical legal studies.

## 2. Introduction

Oral arguments before the U.S. Supreme Court represent one of the most consequential forms of institutional deliberation in American governance, shaping legal outcomes through the most public and visible part of the decision-making process. Within this setting, interruptions by justices are not merely conversational phenomena; they may reflect—and reinforce—underlying power asymmetries among participants. Prior research has shown that female advocates are significantly more likely to be interrupted than male counterparts, even when controlling for experience and institutional position (Cai et al., 2021). While these findings highlight the gendered dynamics of courtroom discourse by quantifying interruption behavior in terms of frequency, less is known about the consequences of such interruptions—specifically, whether they alter the content or emotional tone of an advocate's message.

Therefore, our study extends the original idea to investigate the discursive effects of interruptions during Supreme Court oral arguments, with a particular focus on gender. We pose two primary research questions:

*1. Do interruptions alter or preserve the core semantic content of an advocate's argument?*

*2. Are interruptions directed at female advocates characterized by more negative sentiment than those directed at male counterparts?*

Our aim is twofold: (1) to evaluate whether interruptions disrupt not only the flow but also the substance of advocacy, and (2) to assess whether gendered patterns in rhetorical effect are present in how interruptions are enacted. By combining computational techniques with institutional context, this study contributes to a growing body of work at the intersection of political communication, judicial behavior, and gender studies.

Gender and Discourse | Supreme Court Oral Arguments | Interruptions | Sentiment Analysis | Semantic Similarity | Lexicon-Based Analysis | Computational Linguistics

## 3. Data and Methods

### 3.1. Data Source and Unit of Observation

This study draws on the ConvoKit Supreme Court Corpus, a structured dataset of transcripts before the United States Supreme Court. Our analysis focuses on the 2010–2019 term years, yielding a final analytic sample of 12,663 speech chunks delivered by legal advocates to study the most recent trends.

Each chunk represents a continuous segment of speech by an advocate, bounded by a complete turn in interaction with a single justice. These chunks were created by grouping together advocate utterances that occur between interruptions or pauses in dialogue. Each chunk captures a coherent portion of an advocate's argument — typically a few sentences long — and is treated as a single unit of observation.

We followed the filtering logic established in the original study by Cai et al. (2022), which excludes cases with incomplete metadata, unidentifiable advocate turns, or topics likely to introduce strong confounds in emotional language. In particular, cases involving highly emotive issues such as abortion—identified using a `female_issue` flag—were removed to better isolate gendered interactional dynamics from topic-driven sentiment. Although the original implementation was adapted for our environment, the underlying filtering criteria and rationale remain consistent with the authors' design.

### 3.2. Feature Construction

Interruptions were programmatically identified by detecting sequences in which an advocate's utterance was prematurely cut off and immediately followed by a justice's interjection. In the transcript data, such interruptions are typically marked by double dashes ("—") to indicate an abrupt cutoff or ellipses ("...") to represent trailing speech. These textual patterns were cross-referenced with speaker metadata to confirm that the next speaker was a justice, ensuring that only justice-initiated interruptions were flagged.

The example below shows an interruption exchange between an advocate and Justice Scalia:

**Douglas Laycock (Advocate):**

*"Well, some courts have said yes. There's very little in this record about full beards and whether they're safe or whether they're dangerous... "*

**Justice Scalia (Interrupting):**

*"Mr. Laycock, the problem I have with— with your client's claim... "*

In this exchange, the justice redirects the conversation mid-sentence towards his own analogy, interrupting the advocate's argument. Such interruptions were automatically identified at the utterance level and used to label whether a given speech chunk was interrupted. These interruption labels were critical in structuring both the sentiment and semantic analyses that followed.

## 4. Analysis

### 4.1. Semantic Shifts: Methodology and Implementation

To evaluate whether interruptions alter the semantic content of advocates' arguments, we computed sentence-level semantic embeddings using pre-trained GloVe word vectors (100-dimensional). Each chunk was first preprocessed using standard text-cleaning procedures: conversion to lowercase, removal of punctuation, and expansion of contractions. Tokens were then extracted and filtered to retain only those present in the GloVe vocabulary.

For each token-matched chunk, we calculated a sentence-level embedding by averaging the GloVe vectors corresponding to its valid tokens. This yielded a single vector representation per chunk, capturing its overall semantic content in continuous vector space.

To compare speech under different interruption conditions, we grouped chunks by advocate and interruption status. For each advocate, we aggregated two composite embeddings: one representing their *interrupted chunks*, and one for their *non-interrupted chunks*.

We then calculated the cosine similarity between these two embeddings for each advocate. High cosine similarity values (greater than 0.85) suggest semantic consistency across interruption status, while lower values may indicate that interruptions correspond with shifts in argumentative content.

### 4.2. Gendered Sentiment: Lexicon-Based Analysis

To investigate whether interruptions toward female advocates are more emotionally negative than those toward male advocates, we conducted a lexicon-based sentiment analysis focused on emotional tone. We restricted the analysis to interrupted chunks of speech that had been marked with labels and applied the NRC Emotion Lexicon, which associates English words with specific emotions (e.g., anger, fear, sadness) and sentiment polarity (positive or negative).

Each token in a chunk was matched to its corresponding sentiment label. For each interrupted chunk, we computed:

- The total number of emotion-labeled words,

- The negative emotion ratio: the proportion of emotion-labeled words that associated with negative polarity or negative emotions (anger, fear, disgust, sadness, or negative polarity)

We then aggregated negative sentiment ratios by advocate's gender to assess overall trends. A Welch two-sample t-test was performed to compare group means, followed by an ordinary least squares (OLS) regression model. The regression included control variables for advocate experience, case year, and ideological alignment with the justice, in order to isolate gender as a factor influencing the emotional tone of interrupted speech.

## 5. Results

### 5.1. Semantic Effects of Interruptions

To evaluate whether interruptions alter the semantic content of an advocate's argument, we computed the cosine similarity between the average embeddings of interrupted and uninterrupted speech chunks for each advocate. These sentence-level embeddings serve as vectorized representations of each chunk's semantic content.

The distribution of cosine similarity scores is shown in Figure 1. The scores are tightly clustered between 0.98 and 1.00, with a pronounced concentration above 0.99. The mean cosine similarity is approximately 0.997, indicating an exceptionally high degree of semantic overlap between interrupted and uninterrupted chunks of speech from the same advocate. Full summary statistics are provided in Appendix Table A1.

The near-perfect cosine similarity implies that interruptions do not significantly shift the semantic meaning of advocates' arguments, at least as captured by GloVe-based representations. Rather, interruptions may disrupt the rhetorical flow or perceived coherence without altering substantive content. This finding aligns and supports the prior interpretation that interruptions may be more performative or strategic—serving interactional or hierarchical functions—rather than substantive in altering the content of legal advocacy.

### 5.2. Gendered Sentiment of Interruptions

#### 5.2.1. Descriptive Patterns

To assess whether interruptions directed at female advocates are more negatively toned, we conducted a lexicon-based sentiment analysis using the NRC Emotion Lexicon. We focused on interrupted chunks only and categorized words according to both emotional valence (e.g., anger,

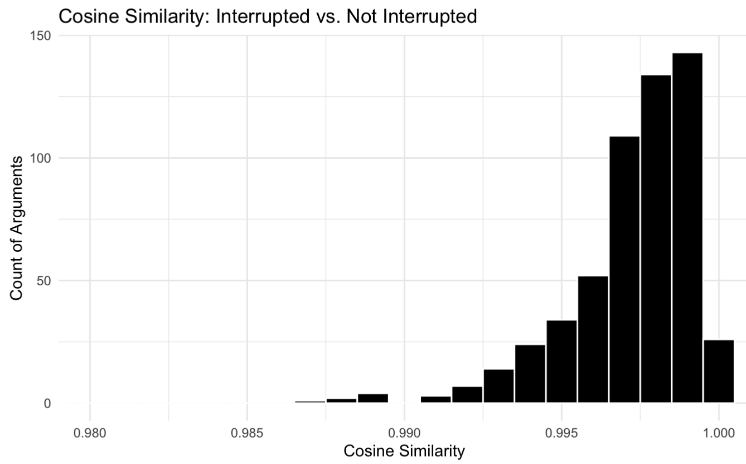



**Fig. 1.** Distribution of Cosine Similarity Between Interrupted and Uninterrupted Speech

fear) and polarity (positive or negative). For each chunk, we computed a negative sentiment ratio, defined as the proportion of emotion-labeled words that were classified as negative.

Figure 2 presents the distribution of negative emotion proportions by advocate gender. Although the overall patterns are broadly similar, interruptions directed at female advocates show slightly higher proportions of anger and disgust, whereas male-directed interruptions exhibit marginally higher levels of fear and sadness. These variations are modest, indicating that emotional tone—when broken down by individual emotion categories—does not differ dramatically by gender.

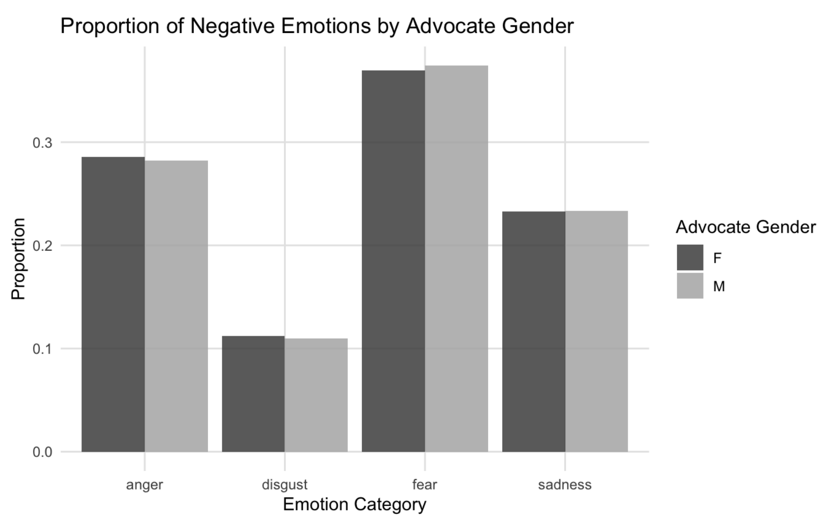



**Fig. 2.** Proportion of Negative Emotions in Interrupted Chunks by Advocate Gender

Figure 3 displays the distribution of negative sentiment ratios across interrupted chunks. Female advocates exhibit a higher median and upper quartile in negative sentiment ratio compared to male advocates. Although the distributions overlap considerably, this pattern suggests that interruptions directed at women tend to be associated with slightly more emotionally negative language. While subtle, this difference is supported by subsequent statistical tests and points to a small but meaningful disparity in the emotional valence of interruptions by gender.

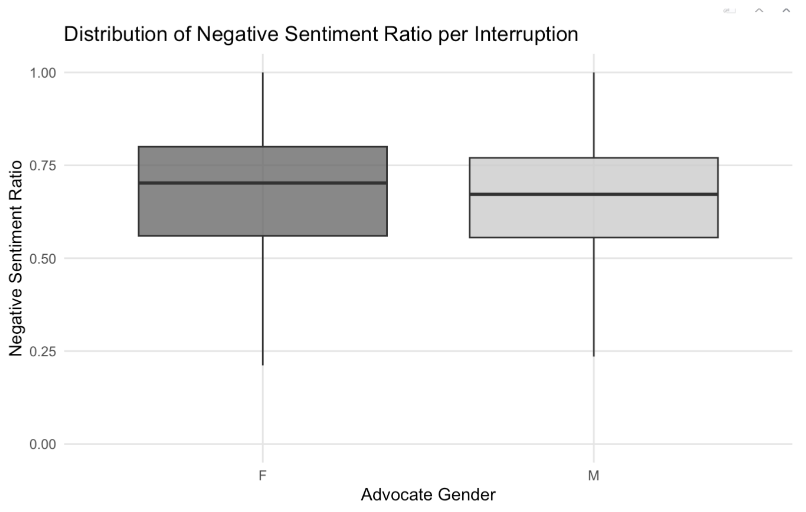



**Fig. 3.** Distribution of Negativity Ratios in Interrupted Chunks by Advocate Gender

#### 5.2.2. Statistical Evidence

Followed, we evaluated gender differences in the emotional tone of interruptions by applying a Welch two-sample t-test and an ordinary least squares (OLS) regression using the negative sentiment ratio as the outcome variable. The Welch t-test reveals a statistically significant difference between the two groups ($p < 0.01$), with female advocates experiencing a higher average proportion of negative emotion in their interrupted speech chunks than males. Although the absolute difference in means is modest with approximately 2%, the result suggests a consistent disparity in how interruptions unfold by gender.

**Table 1.** Key Predictors of Negative Sentiment Ratio in Interruptions

| Variable | Estimate | Std. Error | t value | p-value |
|---|---|---|---|---|
| **Advocate Gender (Male)** | **-0.0114** | 0.0052 | -2.174 | **0.0298** |
| **Advocate Experience (Integer)** | **-0.0022** | 0.0004 | -5.402 | **<0.001 ***** |

Note: Coefficients from OLS regression using negative sentiment ratio as the dependent variable. Gender effect is relative to female advocates.
* $p < .05$, ** $p < .01$, *** $p < .001$

To account for potential confounding variables, we also estimated an OLS regression model controlling for advocate experience, case year, and ideological dynamics between the advocate and the justice. As shown in Appendix Table A2, the effect of gender remains statistically significant: interruptions directed at female advocates are associated with a higher proportion of negative sentiment, even after adjusting for these covariates. At the same time, past experience tempers the negativity of interruptions, while ideological alignment with justices does not offer obvious protection from negative tone.

These findings reinforce the descriptive trends and underscore that gendered differences in the emotional tone of interruptions are not merely artifacts of topic selection or advocate characteristics. While the magnitude of the effect is small, its consistency across statistical approaches adds weight to the interpretation that female advocates face qualitatively different rhetorical dynamics during oral argument.

#### 5.2.3. LDA for Topic Modeling

To explore whether gender differences in emotional tone might reflect underlying differences in the subject matter of advocate speech, we applied Latent Dirichlet Allocation (LDA) topic modeling. Specifically, we implemented a six-topic solution using a document-term matrix constructed from cleaned chunk text. Each chunk was assigned a probability distribution across topics, allowing us to examine whether certain themes were more prevalent among male or female advocates, and whether topic content might explain the observed sentiment patterns.

As a robustness check, we analyzed topic distributions by gender and incorporated topic weights into regression models predicting negative sentiment. average topic proportions were nearly identical across male and female advocates, and the gender coefficient in regression models remained significant after controlling for topic weights. These results indicate that topic content did not account for the gender-based disparities in emotional tone. In addition, top terms across the six topics showed substantial lexical overlap. Words such as court, justice, law, and statute appeared frequently in all topics, highlighting a core limitation of using LDA in this setting:

- Legal discourse is characterized by a highly constrained and repetitive vocabulary, driven by institutional formality and repeated references to statutes and precedent. This reduces topic heterogeneity and undermines LDA's ability to form meaningful distinctions.

- LDA also assumes that documents are sufficiently long to exhibit a mixture of topics. In contrast, individual speech chunks in this corpus tend to remain narrowly focused, often addressing a single legal issue or responding directly to a justice's question.

- For a study of emotion and interruption dynamics, topic groupings must capture differences in legal content—not just "court talk" or standardized courtroom speech.

In sum, while topic modeling was a conceptually valid extension, the constrained language of Supreme Court oral arguments limited its utility. We interpret the LDA results primarily as a reflection of domain-specific challenges, rather than evidence of meaningful topic-based gender divergence.

## 6. Discussion

### *6.1. Conclusions*

This study investigated how interruptions during the U.S. Supreme Court oral arguments affect both the semantic content and emotional tone of advocate speech, with particular attention to gender dynamics. Using GloVe-based embeddings, we found that interrupted and uninterrupted statements from the same advocate had extremely high cosine similarity, suggesting that interruptions do not significantly alter the semantic substance of arguments. However, lexicon-based sentiment analysis revealed that interruptions directed at female advocates were associated with slightly higher negative sentiment ratios, a consistent finding even after controlling for potential confounders such as experience, ideology, and case year.

### *6.2. Theoretical and Empirical Contributions*

This study advances understanding of gendered power dynamics in institutional discourse by shifting focus from the frequency of interruptions to their rhetorical consequences. We assess whether interruptions alter the semantic and emotional content of legal arguments, linking conversational disruptions to broader concerns about epistemic authority and discursive equity and discursive equity in elite legal settings.

Methodologically, we integrate structured transcript metadata with computational textual analysis, applying sentence embeddings and emotion lexicons to Supreme Court oral arguments. Our findings demonstrate both the potential and limitations of using pre-trained language models in formal legal discourse and highlight persistent gendered patterns in how interruptions unfold—even in highly constrained environments, offering a replicable approach for scholars of political communication, judicial behavior, and computational social science.

### *6.3. Limitations and Future Directions*

While our results are suggestive, several limitations remain. First, our sentiment analysis relies on dictionary-based methods that may miss nuanced or sarcastic tones. Second, although interruptions were detected programmatically, our binary categorization (interrupted vs. not) may oversimplify the complex dynamics of turn-taking in oral arguments. Not all interruptions carry the same weight; some may be brief clarifications, while others may sharply derail or redirect an advocate's argument. This dichotomous approach overlooks the variation in interruption severity, tone, and intent.

As a promising direction for future work, researchers could move beyond binary labels to incorporate measures of interruption intensity—such as the frequency built on prior work, or the duration between interruptions. This would enable a more granular analysis of whether frequent or clustered interruptions have compounding rhetorical effects.

Additionally, efforts to disentangle interruption intent using dialogue structure or speech-act classification may help distinguish between adversarial and cooperative interruptions, yielding a more comprehensive understanding of gendered interaction styles. Finally, extending this work through justice-level analysis or cross-institutional comparisons (e.g., lower courts, international tribunals) would broaden its empirical relevance and help contextualize findings within the broader landscape of institutional discourse generalizability.

# A. Appendix

**Table 1:** Welch Two-Sample $t$-Test: Negative Sentiment Ratio by Advocate Gender

| **Statistic** | **Value** | **Notes** |
|---|---|---|
| Mean (Female Advocates) | 0.6614 | |
| Mean (Male Advocates) | 0.6467 | |
| $t$ Statistic | 2.6623 | Welch Two-sample $t$-test |
| Degrees of Freedom | 1729.7 | |
| $p$-value | 0.0078 | ** |
| 95% Confidence Interval | [0.0039, 0.0256] | Difference in group means |

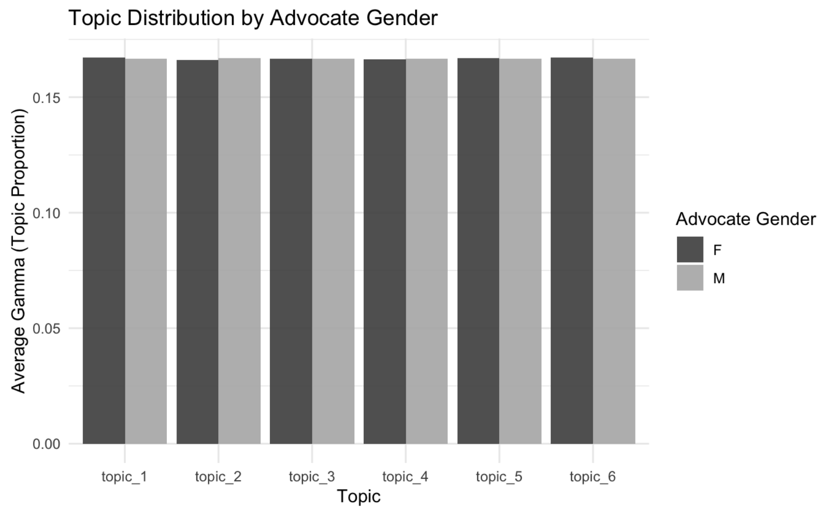
Note: Welch's $t$-test evaluates whether the difference in negative sentiment ratio between female and male advocates is statistically significant.
* $p < .05$, ** $p < .01$, *** $p < .001$

**Table 1.** Welch Two-Sample $t$-Test Results

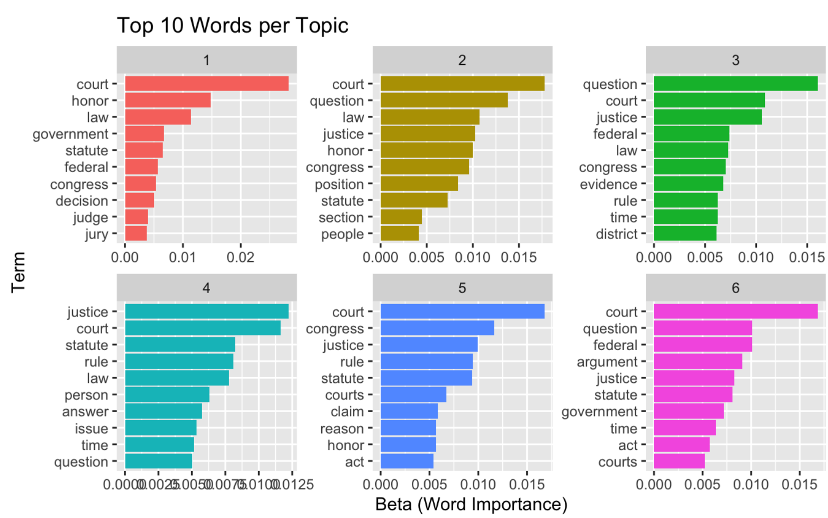
| **Variable** | **Estimate** | **Std. Error** | **p-value** |
|---|---|---|---|
| (Intercept) | -1.4088 | 1.2816 | 0.2717 |
| **Advocate gender (Male)** | -0.0114* | 0.0052 | 0.0298 |
| **Advocate experience (int)** | -0.0022*** | 0.0004 | $< 0.001$ |
| **Case year** | 0.0010 | 0.0006 | 0.1064 |
| Female issue[a] | — | — | — |
| **Advocate ideology (Liberal)** | $0.0061^{\dagger}$ | 0.0035 | 0.0861 |
| Ideology matches | 0.0018 | 0.0036 | 0.6196 |
| Residual Std. Error | 0.1785 (df = 10291) | | |
| Multiple $R^2$ | 0.0041 | | |
| Adjusted $R^2$ | 0.0036 | | |
| F-statistic | 8.501 on 5 and 10291 DF, $p < 0.001$ | | |

$^{*}$ $p < 0.05$ ******* $p < 0.001$ $^{\dagger}$ $p < 0.1$
[a] *Variable dropped due to collinearity or lack of variance.*

**Table 2.** Regression Results: Predicting Negative Sentiment Ratio in Interruptions




**Appendix Figure A1.** Topic Distribution by Advocate Gender




**Appendix Figure A2.** Top 10 Words per Topic

## References

1. Cai, E., Gupta, A., Keith, K. A., O'Connor, B., & Rice, D. ""Let Me Just Interrupt You": Estimating Gender Effects in Supreme Court Oral Arguments." Journal Law Courts, 1–22 (2025) https://doi.org/10.1017/jlc.2024.7
2. Spaeth, H., Epstein, L., Martin, A., Segal, J., Ruger, T., & Benesh, S. "The Supreme Court Database." (2021) http://supremecourtdatabase.org
3. Zhang, J., Spirling, A., & Danescu-Niculescu-Mizil, C. "ConvoKit: A toolkit for the analysis of conversations." Proceedings 20th Conference on Computational Natural Language Learning (2019) https://convokit.cornell.edu
4. Rafo, R., Peña, P., & Emerson, G. "The World Gender Name Dictionary." arXiv preprint arXiv:2112.07656 (2021) https://arxiv.org/abs/2112.07656