EECE 499 - Final Report

Mariam Salman

December 20, 2024

Abstract

This report analyzes a debate transcript using modern language model embeddings to explore semantic patterns and predict future utterances. Dimensionality reduction techniques, including principal component analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), are applied to investigate speaker-specific clustering, temporal trends, and the debate's semantic structure in lower dimensions. Results reveal no distinct low-dimensional patterns separating utterances by speaker or temporal order.

Change-point detection analysis using the *ruptures* PELT algorithm highlights a notable semantic shift along the second principal component (PC2) for one speaker (A) after smoothing, while no significant shifts are observed in the first principal component (PC1). These findings suggest that subtle but meaningful changes in semantic dimensions may require multi-dimensional approaches to detect.

1 Introduction

A central challenge in analyzing multi-speaker debates is predicting future utterances based on prior discourse. This can be conceptualized as:

$$B_{i+1} = f(A_i, C_i, B_i),$$

where B_{i+1} represents the next utterance from a specific speaker (B), and A_i, C_i, and B_i are the most recent utterances from various speakers (A, C, and B).

This study investigates the semantic representation of a debate on the justification of Israel's actions in the Gaza war. The primary objective is to determine whether the embeddings of individual utterances reveal patterns related to speakers, temporal progression, or thematic clustering. Identifying such patterns could inform the development of predictive models for conversational dynamics.

To achieve this, we utilize pre-trained language model embeddings (OpenAI's *text-embedding-ada-002*) to represent utterances as high-dimensional vectors in $\mathbb{R}^{1536}$. Dimensionality reduction techniques are applied to uncover underlying trends:

1. **Principal Component Analysis (PCA):** A linear method to project data onto orthogonal directions of maximum variance.

2. **t-distributed Stochastic Neighbor Embedding (t-SNE):** A non-linear approach that preserves local neighborhood structures in a lower-dimensional manifold.

If distinct clusters or trends emerge, they could guide predictive frameworks. Conversely, the absence of such structures may necessitate alternative approaches, such as topic modeling, change-point detection, or advanced predictive techniques.

2 Data and Preprocessing

2.1 Data Source and Collection

The dataset originates from a YouTube debate on the Gaza war,[1] featuring multiple speakers discussing the justification of Israel's actions. To create a textual dataset suitable for analysis, the video was transcribed using AssemblyAI's[2] speaker diarization service. This process generated a transcript segmented by speaker turns, with each utterance assigned a unique speaker ID (e.g., A, B, C, D, E) and a turn index indicating its sequence in the conversation.

[1] `https://www.youtube.com/watch?v=DQzKw30LeTA`

[2] `https://www.assemblyai.com/`

2.2 Data Structure

Each utterance in the dataset is characterized by:

- **Speaker ID**: A unique label (A, B, C, D, E) assigned by the diarization system.
- **Turn Index**: The sequential position of the utterance within the conversation.
- **Text**: The raw textual content of the utterance as provided by the transcript.

2.3 Preprocessing Steps

Minimal preprocessing was applied to ensure the embeddings accurately represent the semantic content:

- **Basic Cleaning**: Removed extraneous whitespace, corrected transcription errors, and standardized textual elements (e.g., normalized case, removed repeated filler words).
- **Normalization**: Ensured consistent punctuation formatting and removed non-semantic tokens (e.g., hesitation sounds) when present.

3 Methods

3.1 Embedding

Each utterance u is transformed into a high-dimensional embedding vector $E(u)$ using the *text-embedding-ada-002* model:

$$E : U \to \mathbb{R}^{1536},$$

where U represents the set of all utterances. These embeddings capture semantic similarity, such that $\|E(u_1) - E(u_2)\|$ is small when u_1 and u_2 are semantically related.

3.2 Dimensionality Reduction via PCA

Principal Component Analysis (PCA) reduces high-dimensional data to a lower-dimensional subspace while maximizing variance preservation. Given N utterance embeddings arranged in a matrix $X \in \mathbb{R}^{N\times 1536}$, the PCA process involves:

1. Centering the data by subtracting the mean embedding vector.

2. Computing the covariance matrix:

$$\Sigma = \frac{1}{N-1} X^\top X.$$

3. Performing eigen-decomposition on Σ to obtain eigenvectors W and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$.

4. Projecting the data onto the top d principal components:

$$Z = XW_d,$$

   where W_d are the top d eigenvectors corresponding to the largest eigenvalues.

For visualization, we typically select $d = 2$ or $d = 3$.

3.3 Non-Linear Dimensionality Reduction via t-SNE

t-distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction technique that preserves local neighborhood structures. It maps high-dimensional data into a lower-dimensional space by minimizing the divergence:

$$\mathrm{KL}(P\|Q) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}},$$

where $p_{j|i}$ and $q_{j|i}$ represent similarities in high and low dimensions, respectively.

4 Results

4.1 Explained Variance in PCA

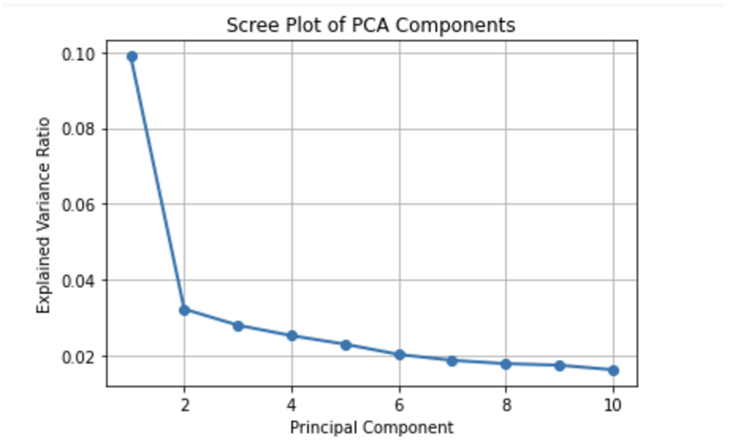


Figure 1: Scree Plot of PCA Components.

Figure 1 shows that the first principal component (PC1) accounts for 9.9% of the variance, followed by PC2 with 3.2%. The remaining components contribute progressively less, indicating a high-dimensional structure without a dominant semantic axis.

4.2 2D PCA Visualization

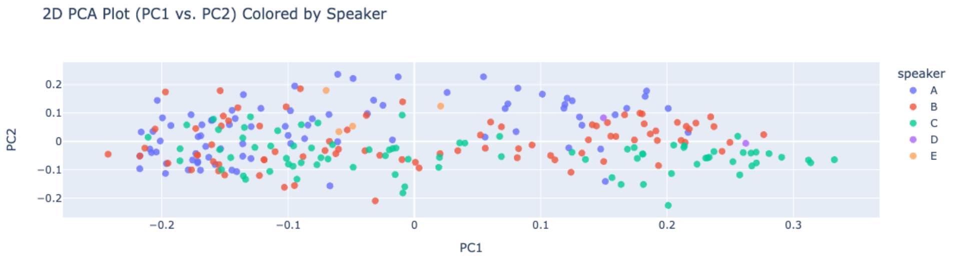


Figure 2: 2D PCA Plot (PC1 vs. PC2) Colored by Speaker.

Figure 2 illustrates that speaker utterances overlap significantly, with no distinct clusters associated with individual speakers in the PC1–PC2 plane.

As shown in Figure 3, utterances do not exhibit a discernable temporal progression. Early and late turns intermix, indicating no clear trajectory over time.

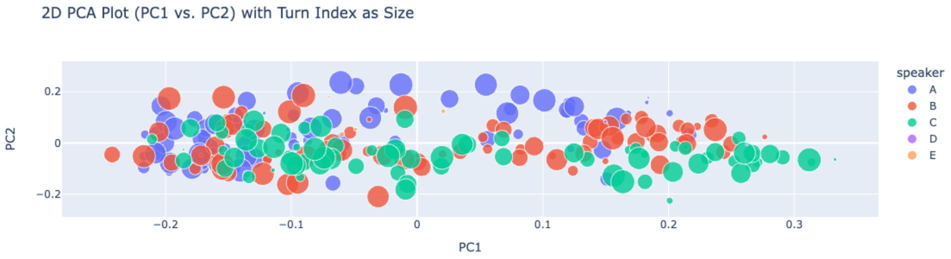


Figure 3: 2D PCA Plot (PC1 vs. PC2) with Turn Index as Marker Size.

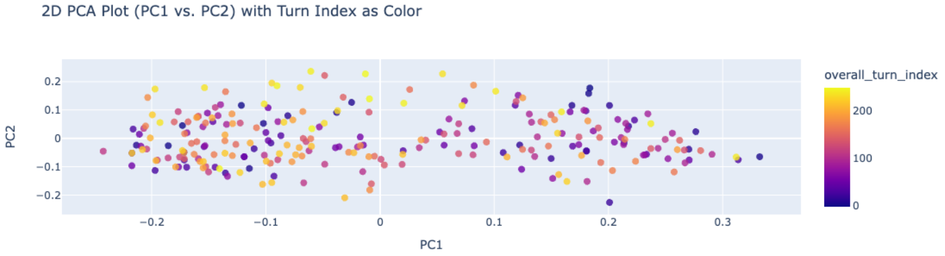


Figure 4: 2D PCA Plot (PC1 vs. PC2) with Turn Index as Color.

Similarly, Figure 4 demonstrates that even when the turn index is encoded as a continuous color scale, utterances from different periods remain interspersed, suggesting no temporal clustering or linear evolution in the semantic space.

4.3 3D PCA Visualization

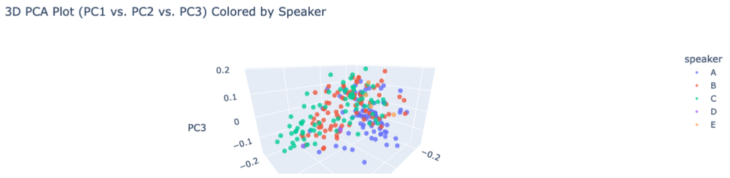


Figure 5: 3D PCA Plot (PC1 vs. PC2 vs. PC3) Colored by Speaker.

The 3D visualization in Figure 5 also reveals no distinct clusters or trajectories. The data forms a diffuse cloud, reinforcing the absence of speaker-specific or temporal patterns in the first three principal components.

5 t-SNE Results and Interpretation

To explore the non-linear structure of the debate's semantic embedding space, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) to the utterance embeddings. Unlike PCA, which focuses on linear variance, t-SNE preserves local neighborhood relationships and can reveal clusters or manifolds that may not be apparent through linear methods.

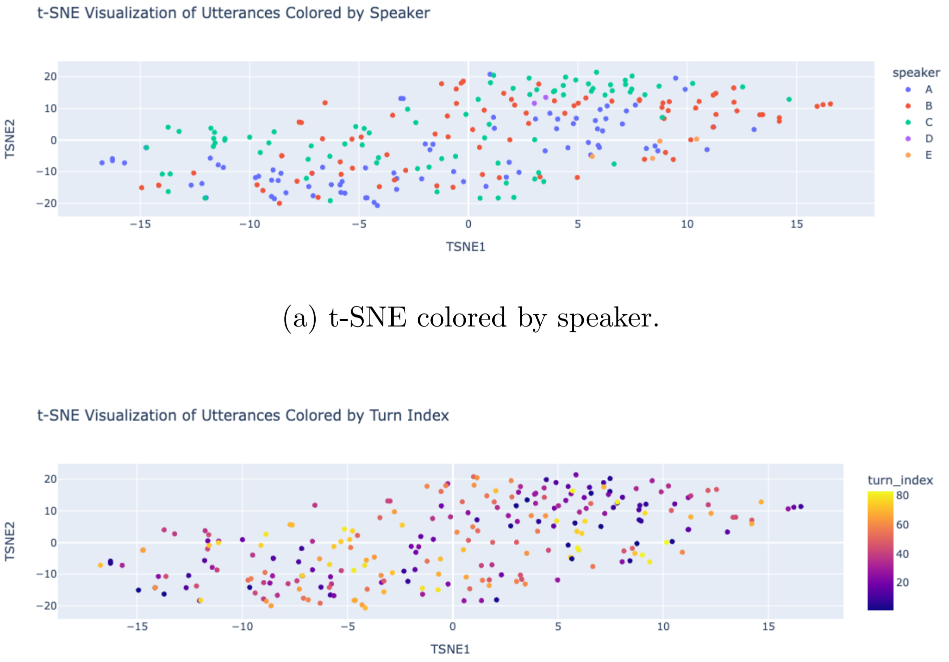
Distribution of Points: The t-SNE visualization (Figure 6a) depicts an elongated cloud of points without distinct, well-separated clusters. Utterances are intermixed, forming a continuous distribution that lacks clear subgroupings. This suggests that the high-dimensional embedding space does not condense into discrete low-dimensional compartments when viewed through t-SNE.

Color Coding by Speaker: When points are color-coded by speaker identity (A, B, C, D, E), no single region of the plot is dominated by utterances from any one speaker. Instead, contributions from each speaker are scattered throughout, indicating that speaker-specific differences are not strongly represented in the embedding space. This suggests that overlapping vocabularies and shared references may blur potential distinctions in speaker styles or rhetorical patterns.

Color Coding by Turn Index: Applying a continuous color gradient to represent the turn index (early to late utterances) reveals no smooth temporal trajectory (Figure 6b). Early and late utterances are intermixed, suggesting that the debate cycles through similar arguments or subtopics rather than evolving linearly. The absence of distinct temporal phases indicates that the conversation meanders through related semantic territories repeatedly.

Theoretical and Comparative Implications: The t-SNE findings, consistent with PCA results, reveal a highly intermixed semantic representation of the debate. Differences in speaker style, stance, or content are not captured as distinct clusters in the embedding space, likely due to the focused nature of the debate topic (the justification of Israel's actions). This homogeneity highlights the limitations of sentence-level embeddings for capturing discourse-level patterns.

These results emphasize the need for more context-aware methods, such as incorporating conversational histories, argumentation structures, or sentiment analysis, to uncover subtle differences. Alternatively, methods tailored to discourse analysis—such as topic modeling or hierarchical representations—could reveal latent structures that are not apparent in simple 2D projections.



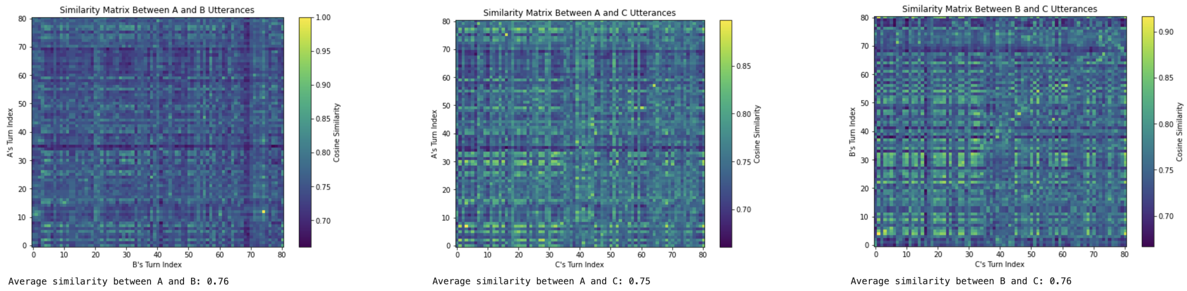
(a) t-SNE colored by speaker.

(b) t-SNE colored by turn index.

Figure 6: t-SNE visualizations of the debate utterances.

6 Integrated Analysis and Observations from Cosine Similarity Matrices

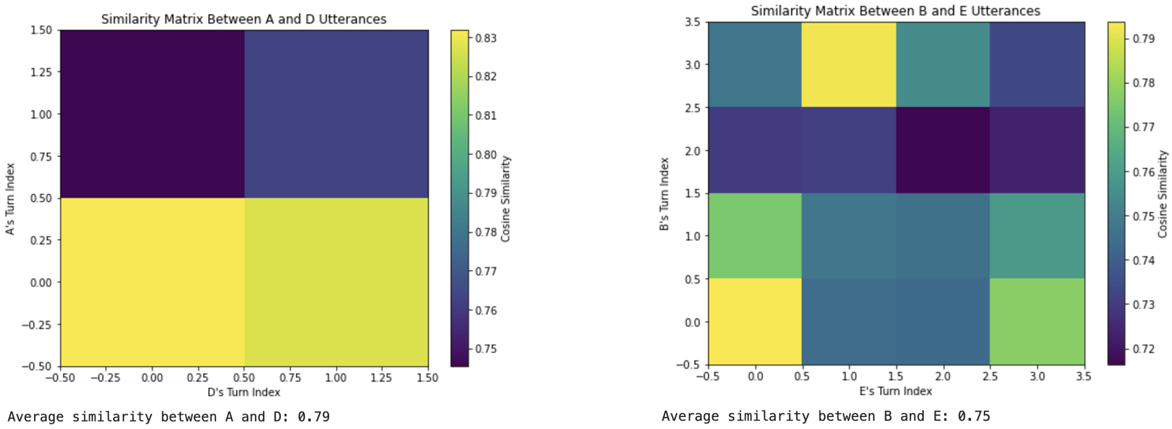
To understand relationships between speakers and compare their semantic content over time, we examined cosine similarity matrices constructed from utterance embeddings. These matrices represent pairwise similarities between speakers, with lighter colors indicating higher similarity.



(a) A vs. B (b) A vs. C (c) B vs. C

Figure 7: Cosine similarity matrices for frequently speaking participants. Matrices show uniform distributions of similarity values without block patterns or diagonal structures indicating temporal or thematic shifts.

General Observations: For speaker pairs with many utterances (e.g., A vs. B, A vs. C, B vs. C), the similarity matrices (Figure 7) exhibit uniform distributions, appearing as “mosaics” with average similarity values between 0.75 and 0.80. This indicates that utterances remain semantically close and consistent, suggesting speakers discuss related concepts throughout the debate.



(a) A vs. D (b) B vs. E

Figure 8: Cosine similarity matrices for less frequent speakers. Similarity remains uniform, reinforcing the homogeneous semantic domain of the debate.

For speaker pairs with fewer utterances (e.g., A vs. D, B vs. E), the matrices (Figure 8) remain uniform and lack clear patterns, further supporting the semantic consistency observed in the larger pairs.

Theoretical Implications: The absence of distinct patterns or segmentation in the similarity matrices suggests that the debate does not progress through identifiable semantic phases. Factors contributing to this uniformity include:

- **Topic Homogeneity:** The debate focuses on a single, narrow issue (Israel's actions in Gaza), limiting semantic diversity and thematic clustering.
- **Sentence-Level Focus:** The embeddings capture sentence-level semantic similarities, potentially overlooking subtle rhetorical or argumentative shifts.
- **Non-Linear Discourse:** Speakers revisit similar arguments, resulting in a consistently intermixed semantic space.

Comparison to Previous Analyses: The similarity matrix findings align with PCA and t-SNE analyses, which also revealed a highly intermixed semantic structure. Together, these results reinforce the conclusion that the debate lacks distinct boundaries or transitions within the embedding space.

Implications for Future Work: The lack of prominent patterns highlights the need for more advanced methods, such as:

- Topic modeling or argument structure analysis to identify finer-grained distinctions.
- Change-point detection applied to topic distributions, sentiment scores, or other features for uncovering subtle shifts.

In summary, the cosine similarity matrices demonstrate that the debate's semantic structure is uniform and continuously interwoven. This calls for nuanced approaches to capture discourse-level phenomena effectively.

7 Change-Point Detection Analysis on Principal Component Scores

7.1 Initial Analysis on PC1

To examine the temporal structure of the debate, we applied the `ruptures` PELT algorithm to detect change points in the PC1 time series for each speaker. Change-point detection identifies indices where statistical properties (e.g., mean, variance) shift significantly:

$$\min_{m} \left[\sum_{j=1}^{m+1} C(x_{\tau_{j-1}+1:\tau_j}) + \beta m \right],$$

where $C(\cdot)$ is a cost function, and β is a penalty parameter.

7.2 Results and Observations

Speakers A, B, C: - Each has approximately 80 turns. - The PC1 time series fluctuates without stable or sustained shifts. - Breakpoints appear at [81] or [83], indicating no internal change points.

Speaker D: - Fewer turns result in a short PC1 time series. - A slight decline in PC1 is observed, but no abrupt shifts occur. - Breakpoints are limited to [2], marking the end of the series.

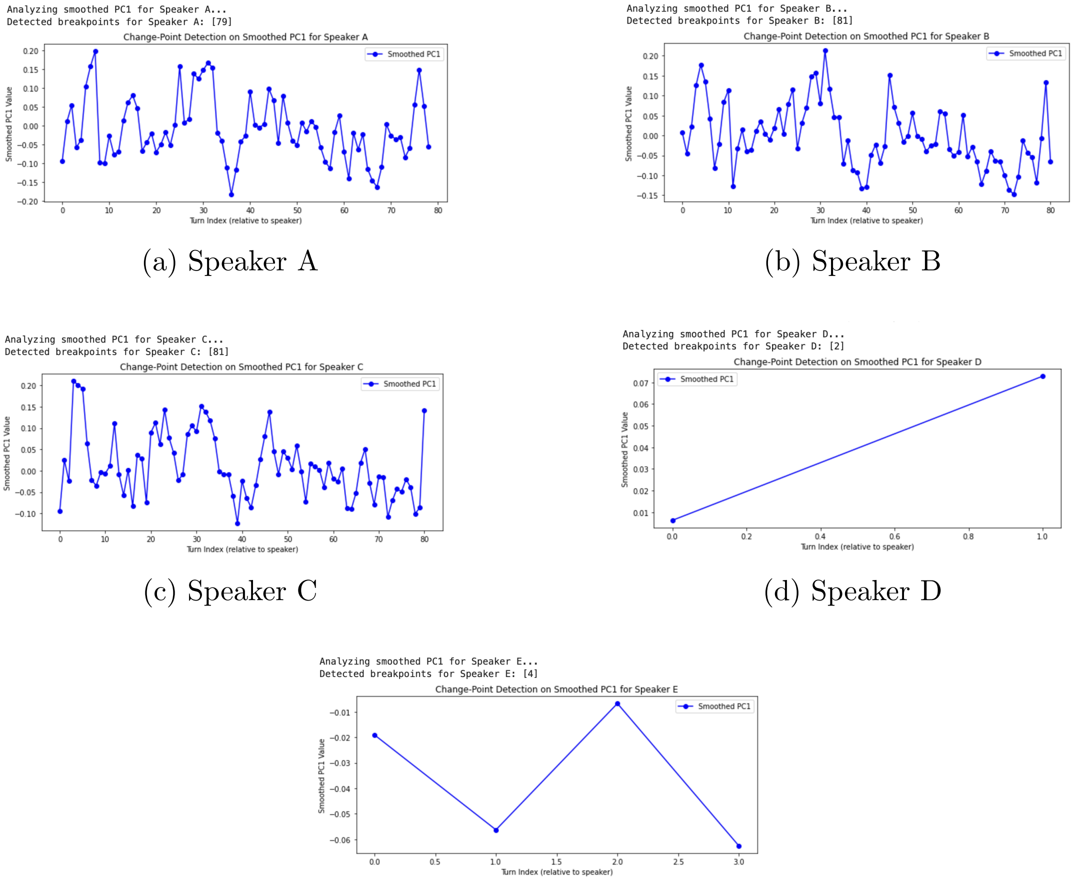
Speaker E: - With only 3–4 points, the data is insufficient to detect meaningful changes. - Breakpoints are [4], indicating no internal splits.

7.3 Extended Experiment: Change-Point Detection on Smoothed PC1 Values

To reduce noise and highlight underlying trends, we smoothed the PC1 time series using a rolling average (e.g., window size = 5). We re-applied the PELT algorithm with a radial basis function (RBF) model to detect subtler, long-term shifts that may not be apparent in raw signals.

7.3.1 Results for Smoothed PC1 Values

General Observations: - Smoothed PC1 trajectories exhibit reduced fluctuations but retain oscillations. - For speakers with many turns (A, B, C), breakpoints still occur only at the final data point, indicating no significant internal changes. - Speakers with fewer turns (D, E) show no meaningful splits, even after smoothing.



```
Analyzing smoothed PC1 for Speaker A...
Detected breakpoints for Speaker A: [79]
```

(a) Speaker A

```
Analyzing smoothed PC1 for Speaker B...
Detected breakpoints for Speaker B: [81]
```

(b) Speaker B

```
Analyzing smoothed PC1 for Speaker C...
Detected breakpoints for Speaker C: [81]
```

(c) Speaker C

```
Analyzing smoothed PC1 for Speaker D...
Detected breakpoints for Speaker D: [2]
```

(d) Speaker D

```
Analyzing smoothed PC1 for Speaker E...
Detected breakpoints for Speaker E: [4]
```

Change-Point Detection on Smoothed PC1 for Speaker E
Smoothed PC1
Smoothed PC1 Value
Turn Index (relative to speaker)

(e) Speaker E

Figure 9: Smoothed PC1 time series for all speakers. Detected breakpoints, if any, are marked.

7.3.2 Theoretical and Practical Implications

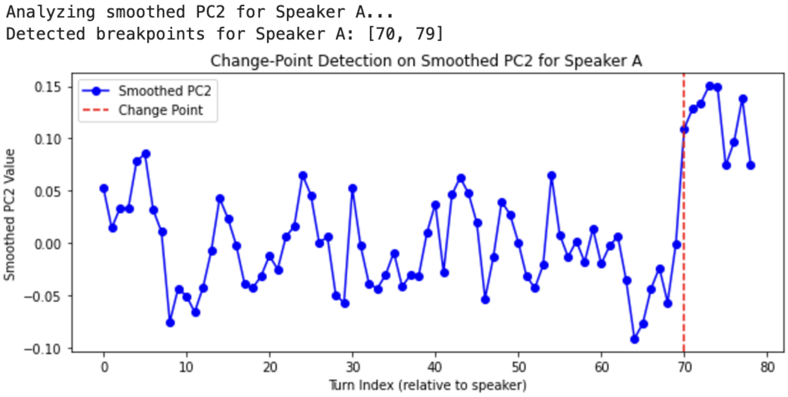
Smoothing did not reveal latent structure in PC1, reinforcing:

- **Semantic Continuity:** PC1 trajectories remain continuous without distinct "chapters."
- **Gradual Changes:** Gradual shifts may not be detected as distinct phases.

- **PC1 Limitations:** PC1 may not align with meaningful thematic boundaries.

7.4 New Experiment: Change-Point Detection on PC2

To capture alternative variance dimensions, we analyzed PC2, the second principal component, using the same smoothing and change-point detection approach. Unlike PC1, PC2 revealed notable results for Speaker A.



```
Analyzing smoothed PC2 for Speaker A...
Detected breakpoints for Speaker A: [70, 79]
```

Figure 10: Smoothed PC2 time series for Speaker A, showing a detected breakpoint.

7.4.1 Results for PC2

For Speaker A, a change point was detected around turn index 70 (e.g., [70, 79]), indicating a significant shift in PC2 values. No similar changes were observed in PC1. For other speakers (e.g., B), only the final breakpoint was detected, consistent with the uniformity seen in PC1.

7.4.2 Interpretation of PC2 Findings

The internal change point for Speaker A suggests that PC2 captures a semantic dimension distinct from PC1. This shift may reflect changes in argumentative style, evidence type, or thematic focus. While most speakers show a uniform landscape across both components, Speaker A's PC2 shift highlights a moment of notable semantic transition.

Implications for Modeling: This result underscores the importance of multi-dimensional approaches in detecting thematic changes. Focusing solely on PC1 risks overlooking critical shifts captured along other axes. Examining utterances near the detected change point for Speaker A could provide insight into the nature of this semantic transition.

8 Discussion

The findings suggest that the semantic space of the debate is complex, with no apparent low-dimensional clustering or temporal patterns in the primary dimension (PC1). Results from t-SNE and cosine similarity analyses reinforce the uniformity of the semantic field. Change-point detection on PC1, even after smoothing, did not reveal internal segments or significant shifts.

In contrast, the PC2 analysis highlights a subtle but meaningful semantic shift for Speaker A, suggesting that not all speakers follow a uniform trajectory. This late-stage shift, captured only in PC2, emphasizes the limitations of focusing solely on a single dimension and underscores the value of multi-dimensional analyses.

These results highlight the need for exploring additional dimensions and incorporating non-linear features, such as topics, sentiment, or argumentative stance, to uncover meaningful patterns and transitions in discourse. A multi-faceted approach is essential for capturing the nuanced dynamics of debates and improving predictive modeling of conversational behavior.

9 Next Steps and Further Analysis

Based on these findings, we propose the following directions for future work:

- **Investigating the Nature of the PC2 Shift:** Analyze Speaker A's utterances around the detected change point to identify specific semantic or rhetorical elements that may have shifted.

- **Exploring Additional Dimensions:** Extend the analysis to higher principal components (e.g., PC3, PC4) or employ non-linear embeddings to capture semantic shifts for other speakers or across the debate.

- **Refining Predictive Models:** Incorporate the identified change point in predictive frameworks by conditioning models on whether utterances occur before or after the shift. This may improve next-utterance predictions and provide insights into semantic transitions.

10 Conclusion

This analysis reveals that while the debate exhibits semantic uniformity in the first principal component and other global representations, subtle but meaningful transitions can emerge along alternative dimensions. The identification of a change point in PC2 for Speaker A highlights the presence of semantic shifts that are not immediately evident in primary dimensions or for all speakers.

These findings emphasize the importance of adopting multi-dimensional and multifaceted approaches to accurately capture the temporal and thematic structure of debates. By leveraging such approaches, future work can enhance the modeling and prediction of conversational dynamics, providing deeper insights into discourse patterns and transitions.