Rhetorical Structure Embeddings: Measuring Rhetorical Organization in Political Texts

Yinxuan Wang*

May 19, 2026

Abstract

Rhetorical organization is politically consequential but difficult to measure at scale. I introduce rhetorical structure embeddings (RSEs), a representation of paragraph-level rhetorical organization derived from adjacent layer transformations in BERT. Combined with an unsupervised dimensionality-reduction and clustering pipeline, the approach discovers recurring organizational patterns in political text. In both a three-source register-recovery task and a content-controlled variant, the RSE outperforms conventional readability, complexity, and grammatical benchmarks, as well as the pooled final-layer output of the BERT model. I apply the method to U.S. Democratic and Republican national party platforms from 1980 to 2024, identifying six recurrent patterns of rhetorical organization and partisan differences in their use. Democrats rely more heavily on a Problem-Solution style, while Republicans more often use a Purpose-Driven style. The method provides a scalable way to discover and analyze rhetorical organization without relying on predefined categories or reducing rhetorical form to a single dimension such as complexity.

[Word Count: 3955]

*PhD Candidate, Wilf Family Department of Politics, New York University. Email: yinxuan.wang@nyu.edu.
I am grateful to Cassandra Handan-Nader, Munroe Kim, Gwyneth McClendon, Bokai Qi, Cyrus Samii, Tara Slough, Joshua Tucker, Luca Vitale, and participants at NYU Methods Workshop, NYU Social Data Science Workshop, and PolMeth 2025 for their valuable comments.

Introduction

How do political elites organize similar commitments into different kinds of appeals? Why do messages advancing similar policies resonate differently with audiences? Answers to these questions do not lie in content alone. They also lie in rhetorical form: how texts package claims, elaborate commitments, and build emphasis. These organizational choices matter politically. They can shape how audiences interpret and evaluate political information (e.g., Bischof and Senninger, 2018, 2025; Blumenau and Lauderdale, 2024; Krishnarajan and Jensen, 2022), and they can reveal how political actors adapt their language across ideological and communicative contexts (e.g., Decadri and Boussalis, 2020; Schoonvelde et al., 2019). More broadly, this distinction between content and form reflects the fact that political language is not simply a neutral vehicle for policy communication, but part of how political action is performed (Finlayson, 2004). Yet despite growing interest in rhetorical form, tools for measuring it systematically at scale remain limited. This paper addresses that gap by proposing an unsupervised representation of paragraph-level rhetorical organization: the way paragraphs sequence and connect claims, elaborations, directives, evidence, and emphasis.

Existing computational approaches to political rhetoric capture several important dimensions of variation, but they do not directly measure rhetorical organization as a distinct object. One family of approaches represents political text through bag-of-words features that abstract away from word order, for example, by using differential word-use rates to capture speaker distinctiveness (e.g., Huang, Perry and Spirling, 2020). These approaches are useful for measuring lexical variation, but they are not designed to capture how content is organized. Missing the organizational side is risky because shared vocabulary can conceal substantial differences in use. As Kraft and Klemmensen (2024) show, actors across ideological camps may use the same moral terms while attaching systematically different meanings to them. A second family of measures, including general complexity and political domain-specific sophistication metrics, captures only a narrow dimension of rhetorical organization (e.g., Benoit, Munger and Spirling, 2019). Two texts can score similarly on conventional complexity metrics but differ sharply in organization, for example, when one relies on declarative statements and the other uses questions to build emphasis. A third family of approaches treats rhetorical forms as predefined categories and learns to predict them with supervised models (e.g., Horák et al., 2024). These approaches depend on predefined labeling schemes and, in some cases, language-specific tools, making them costly to adapt across contexts and less suited to discovering organizational forms that fall outside their training categories.

In this paper, I propose an unsupervised approach to measuring rhetorical organization at the paragraph level. I derive *rhetorical structure embeddings* (RSEs), vector representations that capture paragraph-level organizational patterns, from layer-wise transformations in BERT (Bidirectional Encoder Representations from Transformers). I then use these embeddings in a discovery pipeline that combines dimensionality reduction and clustering to identify recurring rhetorical forms in corpora. To interpret the resulting clusters, I employ a qualitative labeling procedure assisted by large language models (LLMs). Together, the RSEs and the discovery pipeline support typology construction and comparative analysis of rhetorical forms across actors, time, and contexts. I validate the RSE-based discovery pipeline with both a general and a content-controlled register-recovery task in which three text sources overlap in political subject matter but differ in their organizational structure.

The approach draws inspiration from rhetorical structure theory, which treats texts as organized through functional relations connecting spans into coherent structures (Mann and Thompson, 1988). I do not attempt to recover those relations directly, since doing so typically relies on supervised discourse parsers trained on annotated corpora and often tailored to particular languages or annotation schemes. Instead, I adapt the underlying intuition to paragraph-level analysis and represent rhetorical organization through contextual patterns captured by the transformer architecture. The goal is to represent recurring paragraph-level organizational configurations that are realized through, but not reducible to, lexical and syntactic choices.

I apply the method to U.S. national party platforms from 1980 to 2024 and find six recurrent patterns of rhetorical organization. Democrats rely more heavily on a Problem-Solution style that states an issue and then details specific policy remedies, implementation steps, or beneficiary groups, while Republicans more frequently use a Purpose-Driven style that emphasizes collective goals and broad commitments. These differences are consistent with theories of asymmetric politics that characterize Democrats as a coalition of social groups with distinct policy concerns and Republicans as the vehicle of a conservative ideological movement (Grossman and Hopkins, 2016). By providing a scalable measure of rhetorical organization, the paper shows how theoretically meaningful rhetorical strategies can be detected without relying on predefined labels or collapsing rhetoric into a single dimension such as complexity.

Proposed Methodology

Rhetorical Structure Embeddings

I propose a text representation of rhetorical organization derived from pre-trained BERT models. In BERT-base, each token is first mapped to a 768-dimensional vector ($d_{\text{model}} = 768$) by an embedding layer, a learned lookup table that encodes token content and positions. These token vectors are then updated through $L = 12$ transformer layers, which combine information across tokens via self-attention and nonlinear transformations to produce contextualized representations. BERT is well suited to this approach because its bidirectional encoder architecture allows each token to be contextualized by both preceding and following tokens. Its multi-layer architecture produces token representations at each layer, while its open-source implementation makes these intermediate representations accessible for analysis.

Let $\mathbf{Y}_i \in \mathbb{R}^{N \times d_{\text{model}}}$ denote the output of layer i for an input of length N. In BERT, the final step of each layer is a LayerNorm operation, which normalizes each token representation across its feature dimensions and then applies a learned affine transformation with scale γ_i and bias β_i. To improve comparability across layers, I remove this learned affine transformation before computing layer-to-layer differences:

$$\mathbf{Z}_i = \frac{\mathbf{Y}_i - \mathbf{1}_N \beta_i^\top}{\mathbf{1}_N \gamma_i^\top}, \qquad \mathbf{Z}_i \in \mathbb{R}^{N \times d_{\text{model}}},$$

where $\beta_i, \gamma_i \in \mathbb{R}^{d_{\text{model}}}$ are the bias and scale parameters of the final LayerNorm in layer i, extracted from the pre-trained model, $\mathbf{1}_N$ is an N-dimensional vector of ones, and division is applied elementwise.

For adjacent layers $i-1$ and i such that both $\mathbf{Z}_{i-1}$ and $\mathbf{Z}_i$ are defined, I define the organizational signal at layer i as the change between their normalized representations:

$$\Delta\mathbf{Z}_i = \mathbf{Z}_i - \mathbf{Z}_{i-1}, \quad \Delta\mathbf{Z}_i \in \mathbb{R}^{N \times d_{\text{model}}}.$$

I then aggregate token-level signals into a single input-level vector using masked mean pooling. For batch processing, sequences are padded with special padding tokens so that all inputs in a batch have the same length. Let $m_t \in \{0, 1\}$ indicate whether position t corresponds to a real, non-padding token. I compute the mean over non-padding positions to ensure that padding does not affect the pooled representation. For each embedding dimension j,

$$\Delta Z_{i,\text{para}}[j] = \frac{\sum_{t=1}^{N} \Delta\mathbf{Z}_i[t, j] \cdot m_t}{\sum_{t=1}^{N} m_t}, \quad \Delta Z_{i,\text{para}} \in \mathbb{R}^{d_{\text{model}}}.$$

Prior work in "BERTology" finds that syntactic information is concentrated in the middle layers of BERT-base, roughly layers six to nine (Rogers, Kovaleva and Rumshisky, 2020). I therefore construct the paragraph-level RSE from an adjacent-layer transformation within this middle-layer band:

$$\text{Rhetorical Structure Embedding} = \Delta Z_{i^*,\text{para}}, \quad \Delta Z_{i^*,\text{para}} \in \mathbb{R}^{d_{\text{model}}},$$

where i^* indexes a selected middle-layer transformation. In the analyses below, I use the 8–9 transformation, $\Delta Z_{9,\text{para}}$, which lies at the upper end of this range while remaining below the final layers that more strongly emphasize contextual semantics. In the register-recovery validation task, nearby middle-layer transformations perform similarly but slightly worse; Appendix B reports those comparisons.

Unsupervised Pipeline

Figure 1 summarizes the pipeline: light preprocessing, segmentation and truncation, BERT encoding and RSE extraction, dimensionality reduction, clustering, and cluster interpretation. The unit of analysis throughout is the paragraph. The representation-and-clustering pipeline for discovering paragraph-level rhetorical organization is unsupervised, although some hyperparameter choices are informed by tuning diagnostics and substantive considerations, and cluster interpretation requires human judgment.

Before encoding paragraphs with BERT, I apply light preprocessing to standardize the input text. This step removes special symbols such as emojis, normalizes whitespace, and standardizes punctuation to reduce spurious tokenization differences across otherwise similar paragraphs. The goal is to make the text more consistent for tokenization and downstream embedding extraction, while preserving substantive content.

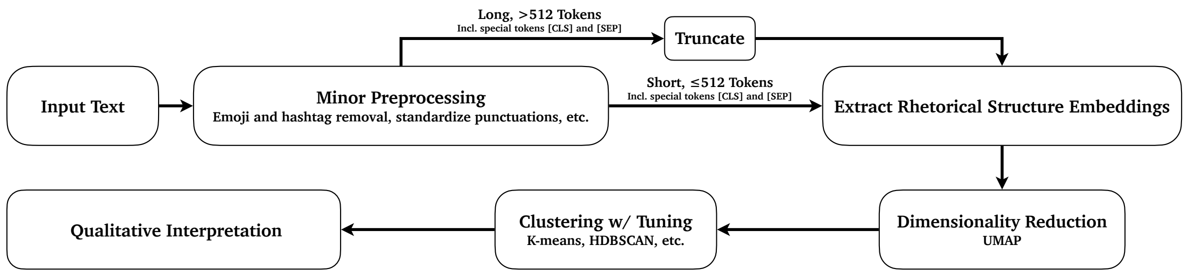
BERT-family models generally restrict inputs to 512 tokens.[1] For paragraphs exceeding the token limit, I use a head-tail truncation strategy that concatenates the beginning and end of the paragraph, allocating more tokens to the tail. This follows evidence from supervised experiments that tail-weighted truncation can preserve predictive signals better than other truncation strategies (Sun et al., 2020).[2] I use extractive truncation rather than hierarchical architectures for long-input encoding because hierarchical designs generally aim at semantic aggregation and are less well suited to preserving the local organizational cues targeted by this method.

Once paragraph-level RSEs are computed, I apply UMAP (Uniform Manifold Approximation and Projection) to reduce dimensionality. UMAP is a nonlinear method that learns a low-dimensional representation that preserves neighborhood structure in the original embedding space

[1]In the paragraph-level corpora used in the validation and application analyses, paragraphs exceeding the 512-token limit are rare. Recent encoder models, such as ModernBERT, support substantially longer context windows (Warner et al., 2024). This study uses standard BERT because the relevant BERTology literature is most developed for the standard architecture and because standard BERT models are widely available in multilingual and language-specific variants. The general strategy of extracting structure signals from intermediate representations could be extended to longer-context encoder models, but doing so would require implementation changes tailored to the model architecture.

[2]The method is also compatible with alternative truncation or extractive compression strategies, so long as the original organizational pattern of the paragraph is preserved as much as possible.

Figure 1: Unsupervised Pipeline Flow Chart



(McInnes et al., 2018). UMAP is widely used in embedding-based workflows because it retains geometry that clustering algorithms exploit while requiring low computational cost (Grootendorst, 2022). In applications, I use a trustworthiness score, which measures how well the reduced neighborhood preserves the high-dimensional structure, to select the number of UMAP components.[3] UMAP emphasizes more local or more global structure through a neighborhood-size hyperparameter. To retain local structure in the embedding space and support clustering that recovers finer-grained organizational patterns, I use relatively small neighborhood sizes in the substantive analysis.

I cluster paragraphs based on proximity in the reduced embedding space after dimensionality reduction. The reduced embeddings are compatible with any standard clustering algorithm. I use two algorithms for different purposes. For the validation task where the number of ground-truth categories is known, I use k-means, which partitions data into k clusters by minimizing within-cluster variance (Lloyd, 1982; MacQueen, 1967). For discovery purposes in the substantive application, I use HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), which identifies dense, stable clusters without assuming spherical shape and labels low-density points as noise (Campello, Moulavi and Sander, 2013; McInnes, Healy and Astels, 2017).

To interpret the clusters, I use a qualitative labeling procedure assisted by an LLM. For each cluster, I randomly sample paragraphs and prompt an LLM to generate summaries of their shared

[3]The formula for the trustworthiness score is provided in Appendix A.

rhetorical features. I then compare those summaries against the sampled texts to ensure that they accurately reflect patterns in the data. Based on this review and direct reading of the cluster samples, I assign concise labels that capture the most prominent recurring rhetorical patterns.

Performance in Register-Recovery Validation

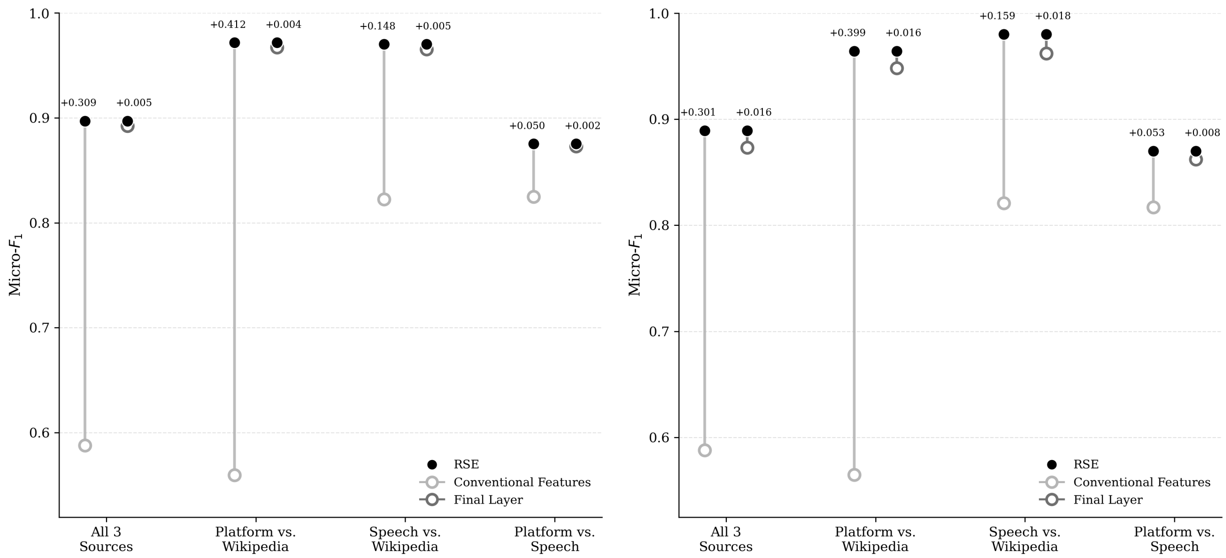
I validate the proposed representation with a three-source register-recovery task. Register recovery is not the primary goal of the method, and the source labels should not be treated as direct descriptions of rhetorical organization. Rather, the task provides a relatively controlled setting in which the three source types are expected to differ in textual organization in ways that the RSE is designed to capture. The validation dataset combines U.S. national party platforms, U.S. presidential campaign speeches, and Wikipedia pages describing presidential candidates' political positions. Party platforms and campaign speeches are collected from the *American Presidency Project*, and Wikipedia texts are drawn from the Wikimedia dump service. I restrict all three sources to the period 2000–2024. Although they overlap in political subject matter, they differ systematically in how that content is organized. Wikipedia paragraphs are descriptive prose narrating factual content, campaign speeches are oral and mobilizational, and party platforms are formal written documents that combine policy description with programmatic and engagement-oriented language.

I run two versions of the task. In the general version, I sample 4,000 paragraphs from each source for a total of 12,000 paragraphs. In the content-controlled version, I restrict all three sources to paragraphs containing terms from a short macroeconomics and jobs dictionary, and then sample 500 paragraphs from each source for a total of 1,500 paragraphs. Both versions use the same pipeline. The RSE is extracted from the layer 8–9 transformation ($\Delta Z_{9,\text{para}}$) of BERT-base-cased, reduced with UMAP to six dimensions, and clustered with k-means ($k = 3$). Because cluster indices are unordered, I match clusters to the known source labels using the Hungarian algorithm (Kuhn, 1955) and evaluate recovery with Micro-F_1, which summarizes classification performance across classes.[4]

Figure 2 shows the gains of the RSE over two benchmarks: a vector of conventional readabil-

[4]Appendices B and A report UMAP tuning details and the Micro-F_1 formula.

Figure 2: Register-Recovery Validation Results



Note: Points show Micro-F_1 scores after optimal label matching of k-means clusters. The RSE uses the layer 8–9 transformation, $\Delta Z_{9,\text{para}}$, of BERT-base-cased; the conventional benchmark uses a concatenated vector of readability, complexity, and grammatical features; and the final-layer benchmark uses the same model's pooled layer-12 output. Panel (a) reports the full validation task, and panel (b) reports the content-controlled task after filtering each source with a short macroeconomics/jobs dictionary. Pairwise comparisons restrict the three-source cluster labels to the two relevant sources without re-clustering.

ity, complexity, and grammatical features, and the pooled final-layer output of the same BERT model. Across both validation tasks, the RSE strongly outperforms the conventional benchmark and also exceeds the final-layer benchmark in every comparison. In the unrestricted three-source task, the RSE achieves a Micro-F_1 of 0.897, compared with 0.588 for the conventional features and 0.892 for the final layer. In the content-controlled task, where all three sources are restricted to overlapping macroeconomics and jobs content, the same ordering holds and the lead relative to the final layer becomes slightly larger. The pairwise comparisons show the same pattern, with positive gains over both benchmarks in every case.

The final-layer comparison is the more demanding benchmark. A pooled layer-12 representation is a strong contextualized summary of paragraph meaning, and semantic information is itself useful for register recovery. The RSE should therefore not be expected to outperform the final layer by a large margin in this setting. The relevant question is whether the RSE captures organizational differences beyond what the final layer does. The RSE outperforms the pooled final layer in every comparison, with a larger gain in the content-controlled task, suggesting that it preserves organizational information more effectively when texts overlap more closely in substantive content. At the same time, this comparison highlights a point of caution in interpretation. In natural language, rhetorical organization is realized through lexical and syntactic patterning, so complete separation from semantic content is neither expected nor required for measuring organizational forms. The claim is not that the RSE isolates a purely structural signal, but that it provides a more organization-centered representation than the pooled final-layer embedding.

Appendix B provides additional details on data construction, dictionary filtering, tuning procedures, and comparisons with alternative layer specifications and benchmark variants.

Rhetorical Organization in U.S. National Party Platforms

National party platforms provide a useful setting for studying rhetorical organization. Platforms not only state positions but also package those positions for voters, activists, and coalition partners, and similar policy content can be organized in different ways. Variation in how platforms organize content is therefore politically meaningful because it can reveal differences in party strategy. Using the proposed RSE, I ask two related questions: what recurring patterns of rhetorical organization appear in U.S. Democratic and Republican national party platforms, and how do the two parties differ in their use of those patterns?

I collect all paragraphs in Democratic and Republican national platforms from 1980 to 2024, comprising 9,371 paragraphs, from *The American Presidency Project*. I extract RSEs from the layer 8–9 transformation, reduce them with UMAP to six dimensions using a neighborhood size of 9, and cluster the reduced embeddings with HDBSCAN using a minimum cluster size of 90 paragraphs. These UMAP hyperparameters aim to retain very local distinctions between platform paragraphs' organizational patterns, and the HDBSCAN hyperparameter aims to rule out idiosyncratic groupings while allowing smaller but cohesive clusters to emerge.

Table 1: Typology of Rhetorical Organization Patterns in U.S. Party Platforms

Cluster Style	Style Description	Representative Example	Shared-Topic Example (Jobs/Opportunity)
I. Purpose-Driven	Collective modal statements ("we will," "we believe") linked to purposes or broad goals; normative and declarative.	"The Republican Party will once again make America Energy Independent, and then Energy Dominant, lowering Energy prices even below the record lows achieved during President Trump's first term." (GOP 2024)	"We advocate incentives for educating, training, and retraining workers for new and better jobsthrough programs like ...as our country surges ahead." (GOP 1988)
II. Elliptical-Directive	Concise, list-like phrasing; often omits main subjects; directives or bullet-points.	"Encourage entrepreneurship and growth by capping the top marginal rate, ending the death tax, and making permanent the Research and Development credit." (GOP 2004)	"Reduce tax rates ... to increase incentives for all Americans and to encourage more savings, investment, output and productivity, and more jobs for Americans;" (Rep 1980)
III. Serial-Parallel	Syntactic parallelism and repetition; sequences of short sentences or phrases that build rhythm and emphasis.	"Our nation is at an inflection point. What kind of America will we be? A land of more freedom, or less freedom? More rights or fewer?..." (Dem 2024)	"The pursuit of opportunity has defined America... This is a land of opportunity. The American Dream is a dream of equal opportunity for all. And the Republican Party is the party of opportunity." (GOP 2012)
IV. Problem-Solution	Problem/issue statement followed by coordinated remedies; often lists beneficiaries and implementation steps.	"We must continue to address the unique readjustment problems of Vietnam veterans by continuing the store-front counseling, vocational training and job placement programs..." (Dem 1980)	"We must provide an opportunity for worker, including those dislocated by changing technologies to adapt to new opportunities... We are committed to ensuring that meaningful job training..." (Dem 1984)
V. Structured Agenda	Dense, formal paragraphs with multiple clauses, conditions, and qualifications; agenda-like sequencing.	"We will pursue our human rights concerns as a necessary part of overall progress... including the possibility of improved, mutually beneficial economic relations between our two countries." (Dem 1980)	"Because American workers have shown that... they can surpass the competition in international trade, we call for the restoration of presidential Trade Promotion Authority. It will ensure ..." (GOP 2012)
VI. Quantitative	Numerical claims embedded in causal or evaluative structure; comparisons, trends, and quantified outcomes.	"This cruelest tax... raged up to 13.3 percent under Carter-Mondale. We have brought it down to about 4 percent and we strive for lower levels..." (GOP 1984)	"Of the 8.5 million Americans still out of work, 40 percent are... Unemployment among teenagers stands at almost 20 percent..." (Dem 1984)

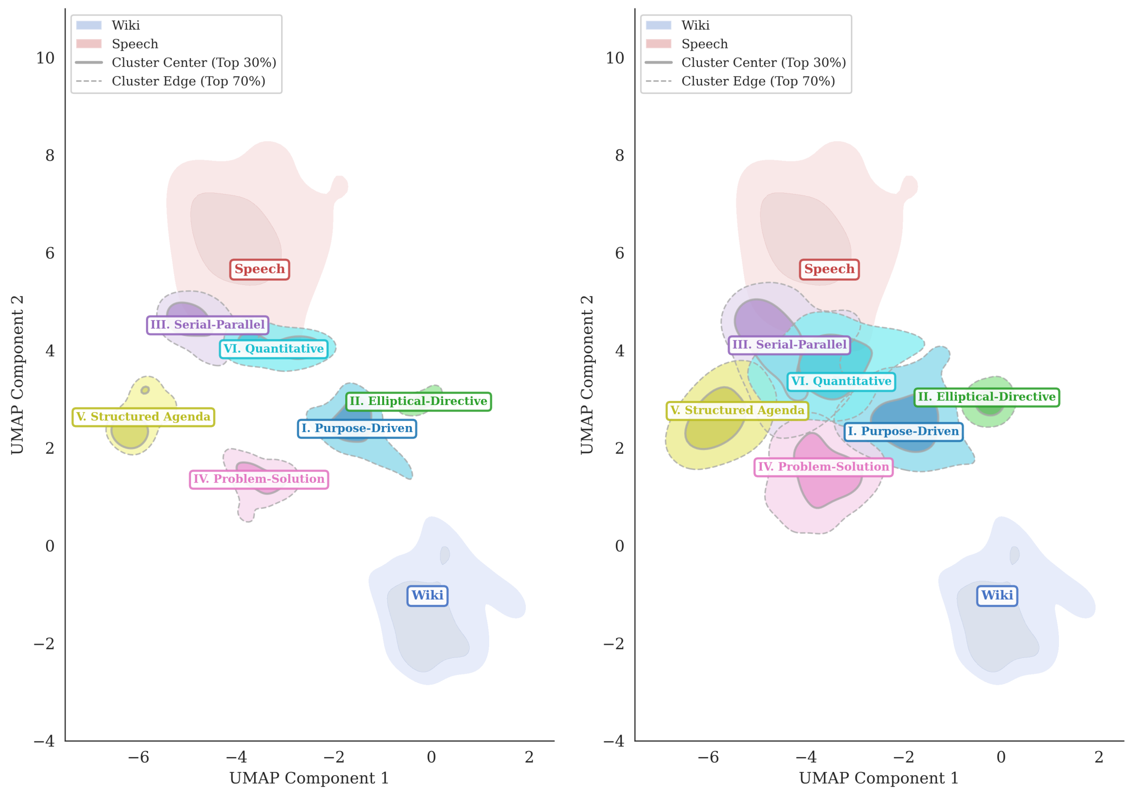
The clustering procedure identifies 2,725 core paragraphs and partitions them into six clusters, with a silhouette score of 0.80, indicating tight and well-separated cluster cores. HDBSCAN labels the remaining paragraphs as noise. For full label coverage, I assign each non-core paragraph to the nearest cluster core in the embedding space; as expected, the silhouette score falls to 0.28 once these more ambiguous cases are included.

To interpret the clusters, I use a qualitative labeling procedure assisted by an LLM. For each cluster, I sample paragraphs from both the core set and the fully assigned set, generate summaries of shared rhetorical patterns using GPT-5.1 with prompts that focus on rhetorical organization rather than policy content, and then audit those summaries through direct reading of the sampled texts. Based on this process, I assign concise labels to each cluster. Appendix C reports prompts and model details.

Table 1 summarizes the six resulting clusters and provides two illustrative examples for each. One is a representative example drawn from any topic. The other is a shared-topic example drawn from jobs- and opportunity-related paragraphs, which provides a convenient basis for cross-cluster comparison within a broad issue area. In what follows, I use *style* as shorthand for these patterns of rhetorical organization. Cluster I (*Purpose-Driven*) consists of declarative, normative sentences linking collective subjects to broad goals. Cluster II (*Elliptical-Directive*) captures short, list-like phrasing that often resembles directives or bullet points. Cluster III (*Serial-Parallel*) relies on repetition and parallelism to build rhythm and emphasis. Cluster IV (*Problem-Solution*) links an issue statement to a sequence of concrete remedies, implementation steps, or named beneficiaries. Cluster V (*Structured Agenda*) consists of dense, formally written paragraphs that pack multiple clauses and conditions into long sentences. Cluster VI (*Quantitative*) embeds numerical claims within broader causal or evaluative arguments.

To visualize cluster separation and locate styles in platform paragraphs relative to other registers, I jointly reduce the platform corpus and the Wikipedia and speech samples used in the validation to two UMAP dimensions and plot kernel-density contours. Figure 3 shows that platform paragraphs occupy an intermediate region between descriptive text (Wikipedia) and engagement-

Figure 3: Platform Rhetorical Styles Relative to Wikipedia and Campaign Speeches

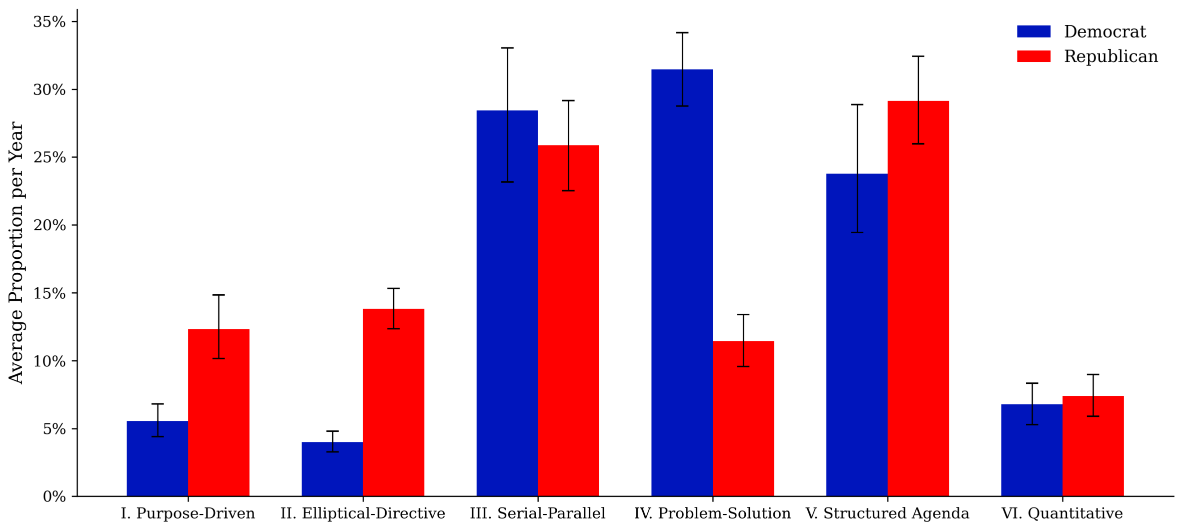


Note: Kernel density estimate contours in 2D UMAP space. Light blue/red show Wikipedia and speech anchors. Other colors show platform clusters. Solid (dashed) contours enclose the 30% (70%) highest-density region. Panel (a): HDBSCAN core points. Panel (b): full corpus after assigning non-core paragraphs to the nearest cluster core.

oriented oral text (campaign speeches), with the Serial-Parallel and Quantitative clusters lying closest to campaign speeches. In the two-dimensional space, the cluster cores are clearly separated, whereas the fully assigned corpus shows more overlap across clusters.

For the following descriptive analysis, I focus on core paragraphs, since they are the most prototypical and best-separated instances of each style. Party usage of each style is calculated from the observed share of core paragraphs. To quantify uncertainty around those shares and the corresponding party differences, I bootstrap paragraphs within each party-year platform document over 5,000 iterations and compute 95% confidence intervals from the resulting distributions. Appendix D reports corresponding estimates for the fully assigned corpus and resampling-based evidence that the core partition is highly stable.

Figure 4: Average Use of Rhetorical Organization Styles by Party

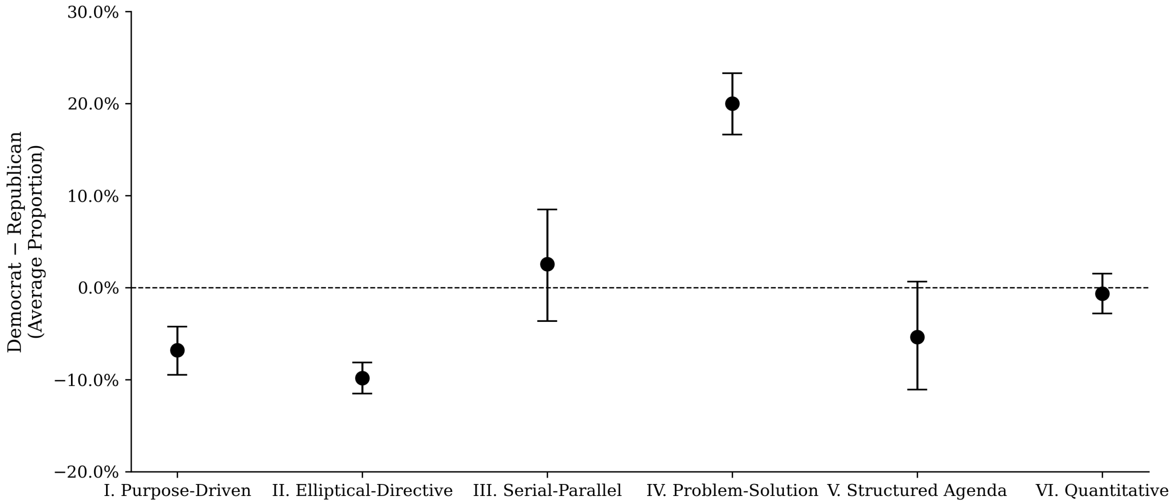


Note: Bars show the average share of core paragraphs in each rhetorical style for Democratic and Republican national party platforms, 1980–2024. Error bars show 95% bootstrap confidence intervals based on 5,000 resamples of paragraphs within each party-year platform document.

Figure 4 shows the yearly average share of core paragraphs in each rhetorical style for Democratic and Republican platforms from 1980 to 2024. Some features of platform writing are common to both parties. Both parties make substantial use of the Serial-Parallel and Structured Agenda styles, suggesting that rhythmic repetition and dense agenda-setting prose are common features of party-platform writing.

At the same time, the overall distributions differ across parties. Democrats rely much more heavily on the Problem-Solution style, while Republicans more often use Purpose-Driven and Elliptical-Directive styles. Figure 5 shows the estimated Democratic-Republican differences in the average shares of all six styles. The bootstrapped 95% confidence intervals indicate that the differences in Problem-Solution, Purpose-Driven, and Elliptical-Directive styles are statistically distinguishable from zero. The contrasts in Problem-Solution and Purpose-Driven are consistent with theories of asymmetric politics that characterize Democrats as a coalition of groups with distinct policy concerns and Republicans as the vehicle of a conservative ideological movement (Grossman and Hopkins, 2016). These descriptive differences in rhetorical organization map closely onto that distinction. Democrats more frequently use a style that states an issue and then details spe-

Figure 5: Democratic-Republican Differences in Average Use of Rhetorical Styles



Note: Points show the mean difference in cluster share between Democratic and Republican platforms (Democratic minus Republican) using core paragraphs only. Error bars show 95% bootstrap confidence intervals based on 5,000 resamples of paragraphs within each party-year platform document.

cific remedies, implementation steps, or beneficiary groups, while Republicans more often use a style that emphasizes collective goals and broad commitments. Appendix D shows that the same patterns persist in the fully assigned corpus.

Discussion

Rhetorical organization shapes how political content is packaged and expressed, but measuring it systematically at scale remains difficult. This paper provides an unsupervised method for representing rhetorical organization at the paragraph level by deriving rhetorical structure embeddings from layer-wise transformations in BERT. Applied to U.S. national party platforms from 1980 to 2024, the method identifies six recurring patterns of rhetorical organization whose usage varies across parties. Democrats rely more on a Problem-Solution style that specifies policy remedies and beneficiary groups, while Republicans more frequently use a Purpose-Driven style that emphasizes collective goals and principles. These patterns describe regularities in how parties present policies and electoral agendas, not causal effects of partisan identity or ideology.

Two points of caution are worth noting. First, rhetorical structure embeddings aim to capture rhetorical organization, but complete separation from semantic content is neither expected nor required, since lexical choice and organization are often intertwined in natural language. The claim is therefore not that the method isolates a purely structural signal, but that it provides a more organization-centered representation than conventional text features or contextualized semantic embeddings. Second, the interpretation process still requires human judgment. LLMs are used for constrained summary generation with fixed sampling rules, prompts, and model versions to support transparency and reproducibility, but manual inspection of sampled paragraphs remains necessary to verify that the summaries accurately reflect each cluster. Interpretation and labeling therefore cannot be fully automated.

Beyond descriptive typology, the method can support several extensions. Rhetorical organization can serve as an outcome variable in studies of party competition, polarization, agenda setting, or electoral context. It can also serve as a predictor, linking changes in rhetorical organization to outcomes such as media uptake, persuasion, or voter learning, especially in designs comparing texts on similar topics. In addition, the identified cluster cores provide a foundation for supervised extensions, since prototypical paragraphs from each cluster can serve as labeled training data for scaling to larger corpora or related domains without repeated clustering.

Taken together, this paper adds a complementary dimension to political text analysis by providing a scalable, unsupervised representation of paragraph-level rhetorical organization. It creates new opportunities for descriptive discovery and theory testing that treat rhetorical organization as a measurable feature of political discourse rather than collapsing it into lexical content, complexity, or predefined rhetorical categories.

AI Disclosure. For rhetorical style cluster interpretation, I used GPT-5.1-2025-11-13 (OpenAI) to generate cluster summaries as part of the interpretation workflow. For manuscript preparation, OpenAI GPT-5.4 and GPT-5.2 were used for copyediting. For coding support, Anthropic Claude Sonnet 3.5 and Claude Opus models, OpenAI Codex 5.4, and OpenAI ChatGPT models, primarily ChatGPT-4o and ChatGPT-5.1, were used for code review, debugging, guided generation of specific code blocks, and code documentation support. All AI-generated outputs were reviewed, edited, and tested where applicable by the author before inclusion.

References

Benoit, Kenneth, Kevin Munger and Arthur Spirling. 2019. "Measuring and Explaining Political Sophistication through Textual Complexity." *American Journal of Political Science* 63(2):491–508. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12423.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12423*

Bischof, Daniel and Roman Senninger. 2018. "Simple politics for the people? Complexity in campaign messages and political knowledge." *European Journal of Political Research* 57(2):473–495. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6765.12235.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6765.12235*

Bischof, Daniel and Roman Senninger. 2025. "Can Simple Language Affect Voters' Political Knowledge and Their Beliefs About Politicians?" *The Journal of Politics* p. 736693.
URL: *https://www.journals.uchicago.edu/doi/10.1086/736693*

Blumenau, Jack and Benjamin E. Lauderdale. 2024. "The Variable Persuasiveness of Political Rhetoric." *American Journal of Political Science* 68(1):255–270. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12703.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12703*

Campello, Ricardo J. G. B., Davoud Moulavi and Joerg Sander. 2013. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*, ed. Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda and Guandong Xu. Berlin, Heidelberg: Springer Berlin Heidelberg pp. 160–172.

Decadri, Silvia and Constantine Boussalis. 2020. "Populism, party membership, and language complexity in the Italian chamber of deputies." *Journal of Elections, Public Opinion and Parties* 30(4):484–503. Publisher: Routledge _eprint: https://doi.org/10.1080/17457289.2019.1593182.
URL: *https://doi.org/10.1080/17457289.2019.1593182*

Finlayson, Alan. 2004. "Political science, political ideas and rhetoric." *Economy and Society* 33(4):528–549.
URL: *http://www.tandfonline.com/doi/abs/10.1080/0308514042000285279*

Grootendorst, Maarten. 2022. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure.".

Grossman, Matt and David A. Hopkins. 2016. *Asymmetric Politics: Ideological Republicans and Group Interest Democrats*. Oxford University Press.
URL: *https://academic.oup.com/book/6600*

Horák, Aleš, Radoslav Sabol, Ondřej Herman and Vít Baisa. 2024. "Recognition of propaganda techniques in newspaper texts: Fusion of content and style analysis." *Expert Systems with Applications* 251:124085. Publisher: Pergamon.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417424009515*

Huang, Leslie, Patrick O. Perry and Arthur Spirling. 2020. "A General Model of Author Style with Application to the UK House of Commons, 19352018." *Political Analysis* 28(3):412–434.
URL: *https://www.cambridge.org/core/journals/political-analysis/article/general-model-of-author-style-with-application-to-the-uk-house-of-commons-19352018/2B1776CD1648E3F4AFBB81D375CCF053*

Kraft, Patrick W. and Robert Klemmensen. 2024. "Lexical Ambiguity in Political Rhetoric: Why Morality Doesn't Fit in a Bag of Words." *British Journal of Political Science* 54(1):201–219.
URL: *https://www.cambridge.org/core/product/identifier/S000712342300008X/type/journal_article*

Krishnarajan, Suthan and Carsten Jensen. 2022. "When Is A Pledge A Pledge?" *British Journal of Political Science* 52(4):1911–1922.
URL: *https://www.cambridge.org/core/product/identifier/S0007123421000284/type/journal_article*

Kuhn, H. W. 1955. "The Hungarian Method for the Assignment Problem." *Naval Research Logistics Quarterly* 2(1-2):83–97.

Lloyd, S. 1982. "Least Squares Quantization in PCM." *IEEE Transactions on Information Theory* 28(2):129–137.

MacQueen, James B. 1967. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. Volume 1: Statistics pp. 281–297.

Mann, William C. and Sandra A. Thompson. 1988. "Rhetorical Structure Theory: Toward a functional theory of text organization." *Text-Interdisciplinary Journal for the Study of Discourse* 8(3):243–281. Publisher: De Gruyter Mouton Section: Text & Talk.
URL: *https://www.degruyterbrill.com/document/doi/10.1515/text.1.1988.8.3.243/html?lang=en&srsltid=AfmBOoqnkCqxDznQZZSHyktBei1RyxrdADbmXNpufpbiXF-RJU0bIS43*

McInnes, Leland, John Healy, Nathaniel Saul and Lukas Groberger. 2018. "UMAP: Uniform Manifold Approximation and Projection." *Journal of Open Source Software* 3(29):861.
URL: *http://joss.theoj.org/papers/10.21105/joss.00861*

McInnes, Leland, John Healy and Steve Astels. 2017. "Hdbscan: Hierarchical Density Based Clustering." *Journal of Open Source Software* 2(11):205.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

Rogers, Anna, Olga Kovaleva and Anna Rumshisky. 2020. "A Primer in BERTology: What We Know About How BERT Works." *Transactions of the Association for Computational Linguistics* 8:842–866.
URL: *https://doi.org/10.1162/tacl_a_00349*

Schoonvelde, Martijn, Anna Brosius, Gijs Schumacher and Bert N. Bakker. 2019. "Liberals lecture, conservatives communicate: Analyzing complexity and ideology in 381,609 political

speeches." *PLOS ONE* 14(2):e0208450. Publisher: Public Library of Science.
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208450*

Sun, Chi, Xipeng Qiu, Yige Xu and Xuanjing Huang. 2020. "How to Fine-Tune BERT for Text Classification?". arXiv:1905.05583 [cs].
URL: *http://arxiv.org/abs/1905.05583*

Warner, Benjamin, Antoine Chaffin, Benjamin Clavi, Orion Weller, Oskar Hallstrm, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, Nathan Cooper, Griffin Adams, Jeremy Howard and Iacopo Poli. 2024. "Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference.". arXiv:2412.13663 [cs].
URL: *http://arxiv.org/abs/2412.13663*

Online Appendix for

Rhetorical Structure Embeddings: Measuring Rhetorical Organization in Political Texts

Table of Contents

A Metrics for Tuning and Evaluation

F_1 Score. F_1 score measures predictive performance of models and is calculated from precision and recall scores. The Micro-F_1 score aggregates true positives (TP_k), false positives (FP_k), and false negatives (FN_k) across classes, so that each observation contributes equally to the final score. Define per-class counts for $k \in \{1, \ldots, K\}$. The micro-F_1 score is

$$\text{Micro-}F_1 = \frac{2\sum_{k=1}^{K} TP_k}{2\sum_{k=1}^{K} TP_k + \sum_{k=1}^{K} FP_k + \sum_{k=1}^{K} FN_k}.$$

In the single-label multi-class setting, $\sum_k FP_k = \sum_k FN_k$ and Micro-F_1 is equal to overall accuracy. I compute Micro-F_1 using the `scikit-learn` implementation (Pedregosa et al., 2011).

Silhouette Score. To summarize cluster coherence and separation in the embedding space, I report the silhouette score. This is descriptive in the platform application since HDBSCAN does not require choosing the number of clusters, but the silhouette score provides a compact measure of within-cluster cohesion relative to separation from the nearest alternative cluster.

For a data point i, the silhouette score is defined as:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)},$$

where a_i is the average distance between i and other points in its assigned cluster, and b_i is the average distance between i and all points in the nearest cluster. The overall silhouette score is the average of all points.

$$\text{Silhouette Score} = \frac{1}{n}\sum_{i=1}^{n} s_i,$$

The silhouette score ranges from -1 to 1, with higher values indicating more distinct clusters. Unless otherwise noted, I compute silhouette scores using Euclidean distance. I compute silhouette scores using Euclidean distance with the `scikit-learn` implementation (Pedregosa et al., 2011).

Trustworthiness Score. To choose the number of UMAP components, I use the trustworthiness score, which assesses how well a low-dimensional representation preserves local neighborhood relationships from the original high-dimensional space. Trustworthiness ranges from 0 to 1, where higher values indicate better preservation of local neighbors. Formally,

$$\text{Trustworthiness}(k) = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i^k} \max(0, (r(i,j) - k))$$

where n is the sample size, k is the number of nearest neighbors evaluated, $\mathcal{N}_i^k$ denotes the set of the k-nearest neighbors of point i in the reduced space, and $r(i,j)$ is the rank of point j among the neighbors of i in the original high-dimensional space. I compute trustworthiness using the `scikit-learn` implementation (Pedregosa et al., 2011).

B Register-Recovery Validation Details and Additional Results

B.1 Data Sources

National Party Platforms. U.S. national party platforms are collected from *The American Presidency Project* (APP) Party Platforms archive. I use Democratic and Republican national platforms and segment each document into paragraphs using the original document formatting. The validation sample is restricted to platforms from 2000–2024.

Campaign Speeches. Campaign speeches are collected from the *American Presidency Project* Campaign Documents archive. To focus on speech-like campaign communication, I filter documents whose titles contain at least one of the following strings: `Rally`, `Speech`, `Address`, `Remarks`, or `Stump Speech`. I then segment documents into paragraphs and remove paragraphs attributed to non-candidate speakers (e.g., audience or crowd responses), retaining paragraphs attributed to the main political speaker.

Wikipedia. Wikipedia texts are extracted from the Wikimedia dump `enwiki-20251101`. Table B1 lists all included pages and extracted sections where applicable. I exclude paragraphs with fewer than 10 words prior to sampling for the validation procedure, which aims to remove headings and fragmentary lines that can appear as paragraph-like units in the dump.

Table B1: Wikipedia Pages Collected

- Views of Lyndon LaRouche and the LaRouche movement
- Political positions of Hillary Clinton
- Political positions of Joe Lieberman
- Political positions of Barack Obama
- Political positions of Rudy Giuliani
- Political positions of John Edwards
- Political positions of the 2008 Republican Party presidential primary candidates
- Political positions of Ron Paul
- Political positions of Mitt Romney
- Healthcare policies of candidates in the 2008 United States presidential election
- Political positions of Joe Biden
- Political positions of Mike Huckabee
- Political positions of Michael Bloomberg
- Environmental activism of Al Gore
- Comparison of the 2008 United States presidential candidates
- Political positions of John McCain
- Political positions of George W. Bush
- Political positions of Newt Gingrich
- Political positions of Herman Cain
- Political positions of Kirsten Gillibrand
- Political positions of Rick Perry
- Political positions of Paul Ryan
- Political positions of Elizabeth Warren
- Political positions of Rand Paul
- Political positions of Ted Cruz
- Political positions of Marco Rubio
- Political positions of Cory Booker
- Political positions of Jeb Bush
- Political positions of Bernie Sanders
- Political positions of Donald Trump
- Political positions of the 2016 Democratic Party presidential primary candidates
- Political positions of the 2016 Republican Party presidential primary candidates
- Political positions of the 2016 United States presidential candidates by political affiliation
- Political positions of Mike Pence
- Political positions of John Delaney
- Political positions of the 2020 Democratic Party presidential primary candidates
- Political positions of Amy Klobuchar
- Political positions of Tulsi Gabbard
- Political positions of Pete Buttigieg
- Political positions of Kamala Harris
- Political positions of Nikki Haley
- Political positions of Ron DeSantis
- Views of Kanye West

B.2 Dictionary of Macroeconomics and Jobs

To construct the content-controlled validation sample, I restrict the dataset to paragraphs that contain at least one term from a short dictionary of macroeconomic and labor-market issues. The macroeconomics portion includes patterns for economy/economic/economics, inflation, recession, interest rates, monetary policy, the Federal Reserve, deficit, debt, budget, tax policy, tariffs, trade, manufacturing, industry, productivity, economic growth, and cost of living. The labor portion includes patterns for jobs, employment, unemployment, workers, labor, workforce, wages, minimum wage, overtime, job training, workforce development, retraining, apprenticeships, labor unions, collective bargaining, working conditions, employee benefits, and unemployment insurance. I apply this filter prior to sampling for the content-controlled validation exercise.

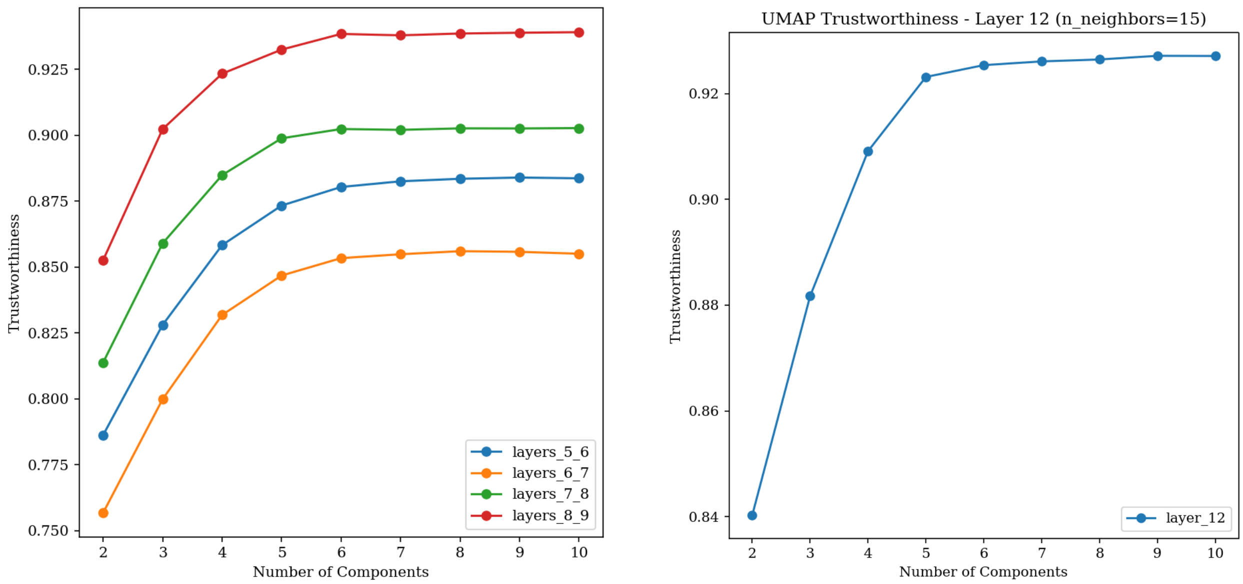
The content-controlled validation sample contains 1,500 paragraphs in total, with 500 paragraphs drawn from each source category. As in the main validation procedure, Wikipedia paragraphs with fewer than 10 words are excluded before sampling.

B.3 Tuning the Number of UMAP Components

I tune the dimensionality of the UMAP projection using the trustworthiness criterion described in Appendix A. For the validation tasks, I use the default UMAP neighborhood size of 15. Figure B1 reports trustworthiness as the number of UMAP components varies from 2 to 10. Panel (a) shows the adjacent layer-pair transformations considered as candidate RSEs in the validation exercise: $\Delta Z_{6,\text{para}}$, $\Delta Z_{7,\text{para}}$, $\Delta Z_{8,\text{para}}$, and $\Delta Z_{9,\text{para}}$. Panel (b) reports the same diagnostic for the pooled layer-12 output used as the final-layer benchmark.

Across these representations, trustworthiness increases sharply as the number of components rises from 2 to 5, while the gains beyond 6 components are small. I therefore use a 6-dimensional UMAP representation for both the RSE and the pooled final-layer benchmark in the validation tasks.

Figure B1: UMAP Trustworthiness by Number of Components



(a) Candidate RSEs

(b) Pooled Final-Layer Output

Note: Panel (a) plots trustworthiness for the adjacent layer-pair transformations considered as candidate RSEs. Panel (b) plots the same diagnostic for the pooled layer-12 output used as the final-layer benchmark.

B.4 Performance under Different Middle-Layer Choice

To assess whether register recovery is stable across adjacent middle-layer transformations, I repeat the three-register validation using layer-to-layer differences ΔZ_6, ΔZ_7, ΔZ_8, and ΔZ_9. For each representation, I run the same UMAP and k-means pipeline as in the register-recovery validation section, using UMAP with 6 components and $n_{\text{neighbors}} = 15$, followed by k-means with $k = 3$.

Table B2 reports Micro-F_1 scores for both the full three-source validation sample and the content-controlled sample. Performance is largely stable across adjacent layer choices within the middle layer range, with ΔZ_9 performing best overall. In the full sample, the three-source Micro-F_1 rises from 0.863 for ΔZ_6 to 0.897 for ΔZ_9. In the content-controlled sample, the corresponding score rises from 0.817 to 0.889. Across both settings, the Platform-Speech contrast remains the most difficult pairwise task, while pairwise recovery involving Wikipedia is consistently high.

Table B2: Register Recovery with Different Adjacent Layer-Pair Transformations

	ΔZ_6	ΔZ_7	ΔZ_8	ΔZ_9
Panel A. Full three-source validation				
All 3 Sources	0.863	0.874	0.877	0.897
Platform vs. Wikipedia	0.956	0.961	0.969	0.972
Speech vs. Wikipedia	0.957	0.956	0.969	0.970
Platform vs. Speech	0.839	0.855	0.848	0.875
Panel B. Content-controlled validation				
All 3 Sources	0.817	0.852	0.869	0.889
Platform vs. Wikipedia	0.928	0.940	0.968	0.964
Speech vs. Wikipedia	0.944	0.958	0.976	0.980
Platform vs. Speech	0.798	0.838	0.835	0.870

Note: Entries are Micro-F_1 scores. Panel A uses the full validation sample of 12,000 paragraphs, with 4,000 paragraphs from each source. Panel B uses the content-controlled sample of 1,500 paragraphs, with 500 paragraphs from each source after dictionary filtering. In both panels, UMAP uses 6 components with $n_{\text{neighbors}} = 15$, and clustering is performed with k-means using $k = 3$.

B.5 Additional Benchmark Results

The main article's validation section presents the main benchmark comparison between the proposed rhetorical structure embedding (RSE), a benchmark based on conventional text features, and the pooled final-layer output of the same BERT model. This appendix provides the corresponding comparison table for the final-layer benchmark and a fuller disaggregation of the conventional-feature benchmark. Throughout, I use the main paragraph-level RSE from the validation analysis, $\Delta Z_{9,\text{para}}$, as the reference specification.

Table B3 reports the conventional-feature benchmark in greater detail. In addition to the concatenated specification using all features jointly, it reports results for individual measures: Flesch Reading Ease (FRE), LIX, mean sentence length, mean word length in characters, and noun density. FRE and LIX are common readability measures. The remaining features overlap with indicators used in prior work on political textual sophistication, although they do not reproduce the domain-specific measure proposed by Benoit, Munger and Spirling (2019).

Table B3: Benchmark Comparison with Conventional Text Metrics

	All	FRE	LIX	Chars /Sent.	Chars /Word	Noun Density
Panel A. Full validation						
All 3 Sources	0.588	0.525	0.560	0.467	0.574	0.556
Platform vs. Wikipedia	0.559	0.526	0.557	0.580	0.589	0.583
Speech vs. Wikipedia	0.823	0.762	0.792	0.734	0.772	0.767
Platform vs. Speech	0.825	0.770	0.783	0.654	0.794	0.752
Panel B. Content-controlled validation						
All 3 Sources	0.588	0.525	0.549	0.481	0.576	0.542
Platform vs. Wikipedia	0.565	0.526	0.554	0.582	0.584	0.594
Speech vs. Wikipedia	0.821	0.762	0.782	0.725	0.780	0.736
Platform vs. Speech	0.817	0.774	0.769	0.644	0.800	0.719

Note: Entries are Micro-F_1 scores. "All" denotes the concatenated benchmark using all conventional readability, complexity, and grammatical features jointly. FRE is Flesch Reading Ease; LIX is the readability index; Chars/Sent. is average characters per sentence; Chars/Word is average characters per word. The full validation task uses random samples from each source. The content-controlled task uses random samples from each source after filtering for paragraphs containing words in a short macroeconomics/jobs dictionary. Pairwise rows restrict the three-way cluster labels to the two relevant sources without re-clustering.

C LLM-Assisted Interpretation Model Version and Prompt

This appendix records the LLM-assisted interpretation setup used for generating initial cluster summaries before human review. I used model `gpt-5.1-2025-11-13` and the following prompt:

```
You are a discourse analyst specializing in rhetorical form,
 ↪  syntax, and political communication.
Your task is to identify the rhetorical structure fingerprint of
 ↪  Cluster {cluster\_id} based on the text samples below.
Focus on form rather than topic. Analyze recurring stylistic
 ↪  patterns in:
- sentence architecture
- clause sequencing
- repetition and parallelism
- contrastive framing
- evidence placement
- discourse progression
- cadence and closure patterns
Only include features that are genuinely recurrent across the
 ↪  samples and distinctive of this cluster.
Exclude:
- topic summaries
- ideological labels
- generic traits common to most political writing
- comments about persuasiveness, emotion, or tone unless tied to
 ↪  a specific structural pattern
Selection rules:
```

```
1. Include only 3-6 features.
2. Each feature must recur across multiple samples, not just one.
3. Prefer distinctive patterns over frequent but generic ones.
4. If the cluster is stylistically mixed or weakly coherent, say
 ↪ so directly.
\# TEXT SAMPLES
[PROVIDE LIST OF SAMPLES]
# OUTPUT FORMAT \hfill
## RHETORICAL STRUCTURAL ANALYSIS
**Cluster Coherence:**
[One sentence stating whether the cluster has a strong,
 ↪ moderate, weak, or mixed rhetorical profile.]
**Dominant Structural Features:**
For each feature, provide:
- **Feature name**
- **Explanation**: one to two sentences describing the recurring
 ↪ structural pattern
- **Coverage**: how many samples show this pattern
- **Example**: one short verbatim example or structural template
 ↪ from the samples
**Style Rationale:**
[A short paragraph explaining how these features combine into a
 ↪ distinctive rhetorical style.]
```

D Rhetorical Patterns in U.S. National Party Platforms, Additional Analyses

This appendix first reports the selection of UMAP dimensionality for the platform application and then presents three additional analyses. First, I reproduce the party comparisons from the main text using the fully assigned corpus, in which paragraphs labeled as HDBSCAN noise are assigned to the nearest cluster core in the reduced embedding space. Second, I assess whether restricting attention to HDBSCAN core paragraphs systematically favors one party or otherwise distorts the substantive comparisons. Third, I report a bootstrap stability diagnostic for the core partition.

D.1 Tuning the Number of UMAP Components for the Platform Application

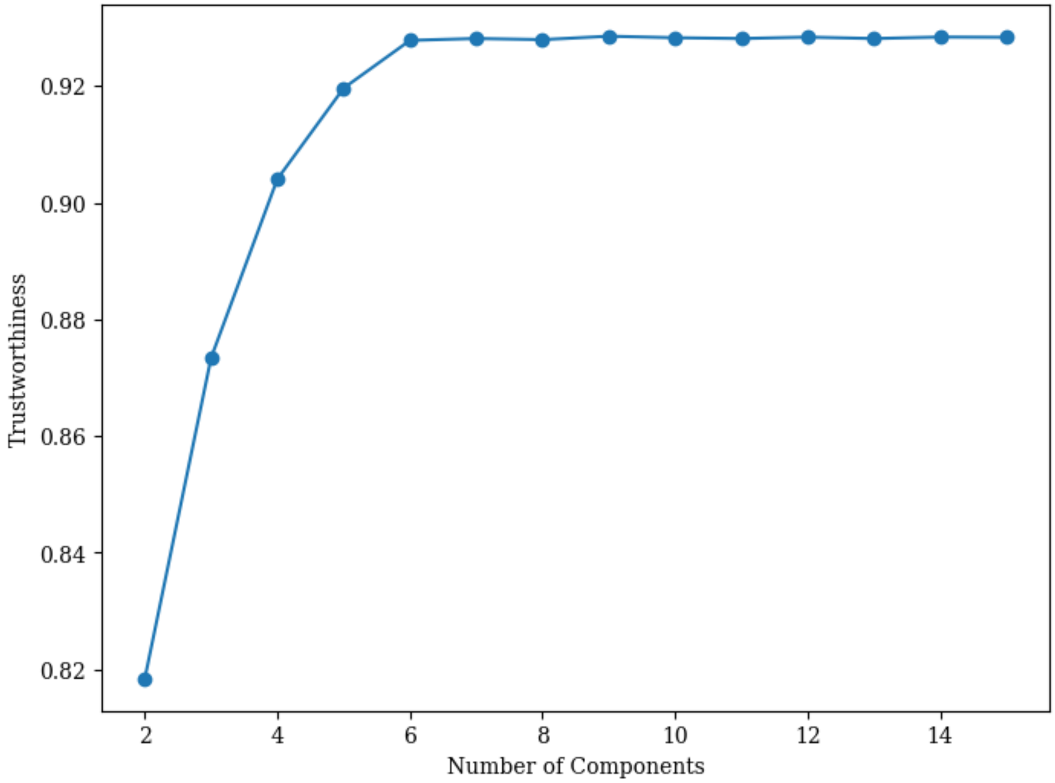
I tune the dimensionality of the UMAP projection for the platform application using the same trustworthiness criterion described in Appendix A. Figure D1 plots trustworthiness as the number of UMAP components varies from 2 to 15 for the paragraph-level RSE used in the application, $\Delta Z_{9,\text{para}}$.

Trustworthiness increases sharply as the number of components rises from 2 to 5, while the gains beyond 6 components are small. I therefore use a 6-dimensional UMAP representation in the platform application. This choice follows the same tuning rule used in the validation analysis and provides a compact reduced space for the subsequent HDBSCAN clustering.

D.2 Results for the Fully Assigned Corpus

Figure D2 reproduces the party-level distribution of rhetorical styles using the fully assigned corpus rather than the HDBSCAN cores alone. The same substantive contrasts appear. Democrats rely much more heavily on Cluster IV (*Problem-Solution*), which accounts for about 43.3% of Democratic paragraphs on average, compared to about 28.5% for Republicans. Republicans continue to use Cluster I (*Purpose-Driven*) more heavily, which accounts for about 18.6% of Republican

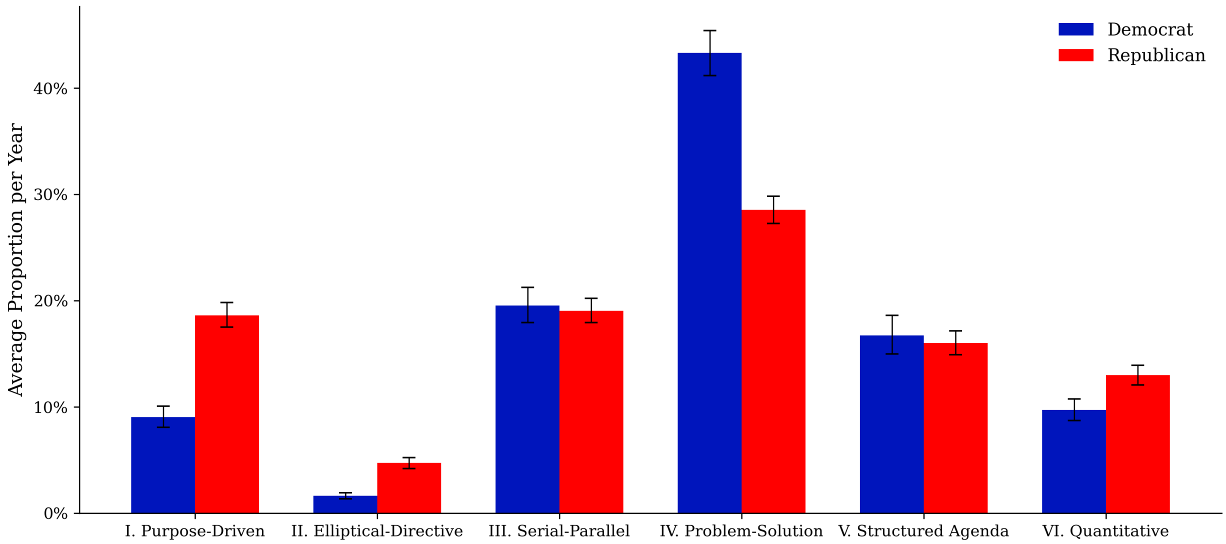
Figure D1: UMAP Trustworthiness by Number of Components, Platform Application



paragraphs on average, compared to about 9.0% for Democrats. At the same time, the fully assigned corpus yields a more even distribution across styles than the core-only analysis, especially for Republicans, because assigning non-core paragraphs to the nearest cluster core brings more low-density and borderline cases into each cluster.

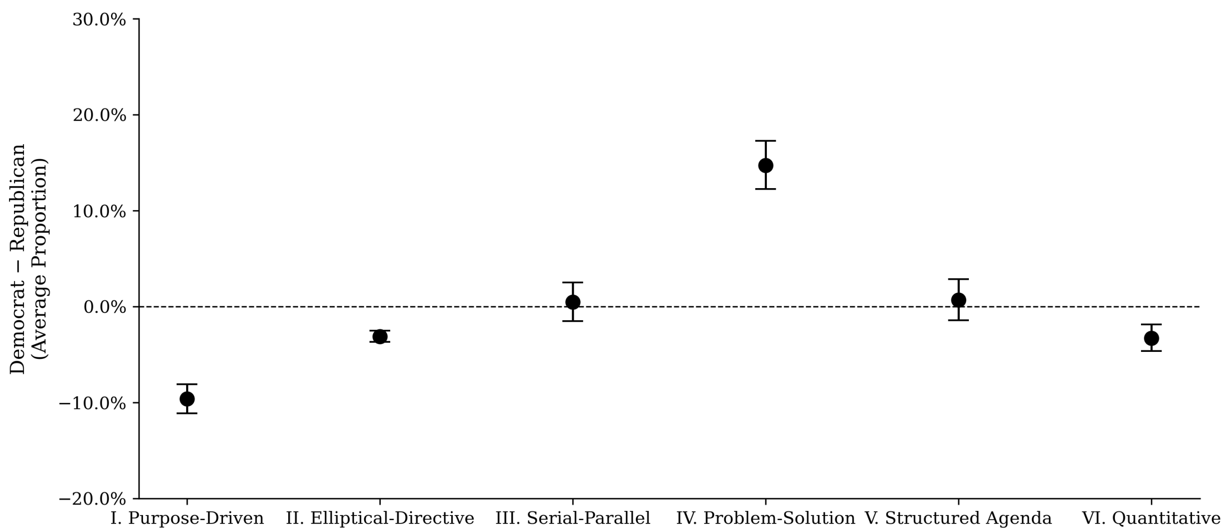
Figure D3 reports the Democratic-minus-Republican difference in the fully assigned corpus. The mean partisan difference is -9.6 percentage points for Cluster I (*Purpose-Driven*) and $+14.8$ percentage points for Cluster IV (*Problem-Solution*), with bootstrapped 95% confidence intervals that do not cover zero in either case. The substantive contrasts highlighted in the main text therefore do not depend on discarding HDBSCAN non-core points.

Figure D2: Average Use of Rhetorical Organization Styles by Party, Fully Assigned Corpus



Notes: Bars show the average share of paragraphs in each rhetorical style for Democratic and Republican national party platforms, 1980–2024, using the fully assigned corpus. Error bars show 95% bootstrap confidence intervals based on the same procedure as in the main text: in each iteration, paragraphs are resampled with replacement within each party-year platform document, cluster shares are recomputed, and confidence intervals are taken from the resulting bootstrap distribution.

Figure D3: Democratic-Republican Differences in Average Use of Rhetorical Styles



Notes: Points show the mean difference in cluster share between Democratic and Republican platforms (Democratic minus Republican) using the fully assigned corpus. Error bars show 95% bootstrap confidence intervals based on resampling paragraphs within each party-year platform document. The dashed horizontal line marks no partisan difference.

D.3 Core-Point Selection by Party and Year

Because the main text focuses on HDBSCAN core paragraphs, a natural concern is whether core selection is systematically related to party. Table D1 reports the proportion of paragraphs retained as cluster cores overall and by party-year. Overall, 30.8% of Democratic paragraphs and 27.8% of Republican paragraphs are classified as core. The year-by-year rates vary substantially across cycles, and they do not reveal a uniform partisan skew: Democrats have the higher core rate in seven platform cycles and Republicans in five. A pooled paragraph-level chi-square test of party by core assignment rejects independence ($\chi^2 = 9.75, df = 1, p = 0.0018$), but this diagnostic treats paragraphs as independent observations and does not adjust for year composition. It should therefore be interpreted descriptively rather than as formal inferential evidence.

Table D1: Core-Paragraph Rates by Party and Platform Year

	Democrat		Republican	
Year	Paragraphs	Core rate	Paragraphs	Core rate
Overall	3,964	30.8%	5,407	27.8%
1980	863	33.6%	565	31.3%
1984	623	25.0%	501	26.7%
1988	30	10.0%	862	34.2%
1992	98	38.8%	456	25.0%
1996	166	25.3%	432	31.3%
2000	263	29.7%	492	35.6%
2004	234	34.6%	600	36.7%
2008	248	39.1%	354	22.9%
2012	260	28.1%	304	22.4%
2016	270	38.9%	358	13.1%
2020	443	36.1%	368	12.5%
2024	466	21.0%	115	10.4%

Notes: The core rate is the share of paragraphs assigned by HDBSCAN to one of the six cluster cores rather than labeled as noise.

A more substantively relevant check is whether restricting the analysis to core paragraphs changes the relative cluster mix within each party. Table D2 compares the aggregate cluster composition of the full and core partitions separately for Democrats and Republicans. HDBSCAN cores systematically overrepresent the more compact, formulaic styles, especially Cluster II (*Elliptical-*

Directive) and Cluster V (*Structured Agenda*), while underrepresenting the more diffuse Cluster IV (*Problem-Solution*) in both parties. However, the sign of the key partisan contrasts does not reverse across partitions. Democrats remain more concentrated in Cluster IV and Republicans remain more concentrated in Cluster I in both the full and core views.

Table D2: Party-Specific Cluster Composition in the Full and Core Partitions

	Democrat		Republican	
Cluster	Full	Core	Full	Core
I. Purpose-Driven	14.2%	9.7%	19.1%	15.2%
II. Elliptical-Directive	3.7%	9.6%	5.6%	18.9%
III. Serial-Parallel	16.9%	19.9%	19.2%	20.1%
IV. Problem-Solution	38.4%	31.4%	26.7%	10.1%
V. Structured Agenda	15.5%	22.1%	15.9%	28.5%
VI. Quantitative	11.3%	7.2%	13.5%	7.1%

Notes: Entries are aggregate paragraph shares within each party, computed over all platform years. "Full" assigns every paragraph to a cluster; "Core" includes only HDBSCAN core points. These are diagnostic totals and therefore differ from the year-weighted bootstrap averages plotted in Figures 4 and D2.

D.4 Bootstrap Stability of the Core Partition

Finally, I assess the stability of the core partition under resampling. In each of 1,000 bootstrap replications, I resample the core paragraphs with replacement, apply the same UMAP reduction used in the application, cluster the resampled observations with k-means using $k = 6$, and compare the recovered grouping to the original core labels. I use adjusted mutual information (AMI) as the primary measure of overlap because it evaluates agreement in grouping while adjusting for chance. This exercise should be interpreted as a recoverability diagnostic for the identified core geometry rather than as a full re-estimation of the HDBSCAN solution.

The bootstrap results suggest the core partition is highly stable. The mean AMI is 0.895, with a 95% interval of $[0.820, 0.945]$. These values indicate that the core grouping is highly recoverable under resampling. I do not treat analogous full-corpus overlap scores as a stability benchmark, because the exhaustive nearest-core assignment deliberately forces low-density boundary paragraphs into clusters. Lower overlap in that setting is therefore expected and mainly reflects the weaker separation of the full partition rather than instability of the identified core geometry.

These appendix analyses support the interpretation in the main text. The fully assigned corpus reproduces the same directional partisan contrasts as the core-only analysis, core selection is not driven by a uniform partisan imbalance across platform cycles, and the core partition itself is highly stable under resampling. The main text therefore focuses on core paragraphs because they provide the clearest and most coherent instances of each rhetorical style, not because the substantive pattern depends on excluding the remainder of the corpus.