Who Counts as "the people"? Measuring Contested Political Concepts from Political Speech using Masked Language Models

Anonymous ACL submission

Abstract

What do political elites mean when they invoke "the people", and how has this meaning shifted across centuries of democratic history? Although political theorists have long argued that "the people" is a discursively constructed and contested category, empirical work has been largely qualitative or limited in scale, and computational approaches have yet to address how contested concepts themselves are framed over time. We address this gap with a generalizable framework that uses masked language models to measure contested political concepts as competitions among theoretically grounded frames: for each occurrence of a target term, we estimate the relative likelihood of theory-guided probe words filling the masked position, yielding interpretable scores without supervision or annotation. We showcase this framework through an in-depth case study of 3,402 U.S. presidential speeches from a period spanning over two centuries (1789–2026), operationalizing "the people" along three dimensions—Nation, Civic, and Global. The resulting measurements align with major historical junctures (the Civil War, the World Wars, the Civil Rights era, Trump's nationalist turn) and covary with indicators of democratic development and macro-economic conditions in theoretically expected directions. Beyond "the people", the framework extends to other contested political concepts.

1 Introduction

Across democracies worldwide, the question of who counts as "the people" has returned to the center of political conflict. Resurgent ethno-nationalism, anti-immigration movements, and exclusionary populism are redrawing the boundaries of political membership in ways unseen for generations. Yet political elites rarely define "the people" explicitly; instead, as opinion leaders shaping mass political cognition (Zaller, 1992), they invoke the term in ways that can quietly reshape its meaning—bounding it by ancestry, by citizenship, or by appeals to a broader human community. That is, these rhetorical shifts are not only descriptive but potentially constitutive: they reshape how citizens perceive the boundaries of their political community and how they treat those they recognize—or refuse to recognize—as civic equals (Anderson, 2020; Smith, 2003; Schmidt, 2008; Stuckey, 2023). Tracking these shifts at scale is therefore a precondition for understanding how democratic membership is reconfigured under political pressure.

Yet despite its theoretical centrality, empirical study on "the people" remains strikingly limited. Qualitative scholarship has documented how national identity is constructed in rich detail (Beasley, 2011; Stuckey, 2023), but has not traced these shifts systematically at scale. Existing computational tools are also poorly suited to the task: unsupervised topic models capture what is discussed rather than how it is framed; supervised approaches, including the fine-tuning of pre-trained models, depend on costly annotation; and zero- and few-shot LLM classification comes at the expense of interpretability and reproducibility. Although computational work has advanced the broader study of political rhetoric (Card et al., 2022; Rule et al., 2015), it has yet to address how contested political concepts themselves are framed over time. A method that captures the competition among multiple frames over a single contested term—at scale, across long historical spans, and without supervision—has yet to be established.

We address this gap with a generalizable framework that measures contested political constructs, such as "the people", as competitions among multiple theoretically grounded frames. For each occurrence of a target term, we replace it with a [MASK] token and use a domain-adapted masked language model (Devlin et al., 2019) to estimate how likely theoretically guided probe words are to

fill masked position, yielding a continuous association ratio across frames without supervision, annotation, or generative inference. For each occurence of "the people", we draw on Kohn's distinction between ethnic and civic nationalism (Kohn, 1944) and on cosmopolitan conceptions of moral community (Nussbaum, 1996) to define three frames: *Nation*, grounded in shared ancestry and bounded sovereignty; *Civic*, grounded in rights-bearing citizenship and constitutional principles; and *Global*, grounded in universalist obligations that transcend state boundaries.

We showcase the framework through an in-depth case study of U.S. presidential rhetoric from 1789 to 2026 (3,402 speeches, 29,424 contextual instances) spanning more than two centuries.[1] Three findings together establish the framework's validity and offer rich substantive insights. First, Nation framing dominated nineteenth-century rhetoric (peaking near 85%) before Civic framing overtook it in the post-Cold War era, followed by a renewed nationalist turn since 2016. Second, the framing shifts align closely with major historical turning points—the Civil War, WWI, the Civil Rights era, and 9/11—which provides strong face validity. Third, regression analyses show that the framing ratios move together with indicators of democratic development and macro-economic conditions in the directions theory predicts.

Our contributions are twofold. Methodologically, we propose a generalizable framework that measures contested political constructs as multi-frame competitions, combining masked language modeling with theoretically guided dictionaries; the framework requires neither supervision nor annotation and yields interpretable, reproducible association scores at the level of individual term occurrences. Substantively, we present the first large-scale longitudinal measurement of how "the people" has been constructed in political speech. Focusing on U.S. presidential rhetoric from 1789 to 2025, we show that the framework-derived measurements align with major historical inflection points and covary with macro-political and economic conditions in theoretically expected directions. Beyond "the people", the framework extends naturally to other contested constructs such as democracy, freedom, and security, enabling the scalable longitudinal study of how core political concepts are reshaped over time.

2 Data Collection and Preprocessing

2.1 Data Collection

To address our research questions, we constructed a dataset of presidential speeches by combining two sources: 1,057 speeches from the Miller Center (`https://millercenter.org/`) and speeches from the American Presidency Project (`https://www.presidency.ucsb.edu/`), an archive of U.S. presidential speeches. From the latter, we retained only entries classified as *Address* and excluded those that could introduce noise into the corpus, such as remarks delivered at informal events and presidential signing statements; this left only speeches directed at the general American public, yielding 2,345 speeches. Both sources were accessed in March 2026 and cover presidential speeches delivered between 1789 and 2026, the full available period. To remove duplicate entries across the two sources, we computed pairwise semantic similarity using sentence embeddings and discarded any speech with a similarity score of 0.99 or higher relative to another document in the corpus. This deduplication step yielded a final corpus of 3,402 unique presidential speeches.

2.2 Model

We employed RooseBERT-cased (Dore et al., 2025) to capture the semantic properties of presidential speech. RooseBERT-cased is a domain-specific pre-trained language model built on the BERT architecture and trained on English-language corpora of political debate and speech, including U.S. presidential debates, parliamentary debates from nine countries, and United Nations proceedings.[2] We applied it directly for contextual embedding extraction and the subsequent analyses, without any additional fine-tuning.

As an initial check that the RooseBERT embeddings capture meaningful structure in our corpus, we applied BERTopic to sentence embeddings projected via UMAP (Grootendorst, 2022). The re-

[1] The U.S. offers a suitable context with a democratic history spanning over two centuries, a multi-cultural polity in which the boundaries of "the people" have been repeatedly redrawn, and a continuous archive of presidential rhetoric that enables longitudinal measurement at otherwise unavailable scale.

[2] The developers of RooseBERT explicitly identify the analysis of political discourse and its influence on public opinion as a primary motivation for the model (Dore et al., 2025), which aligns directly with the goals of our study. RooseBERT-cased is distributed by the Inria Centre at Université Côte d'Azur for non-commercial academic research, and our use is consistent with these terms.

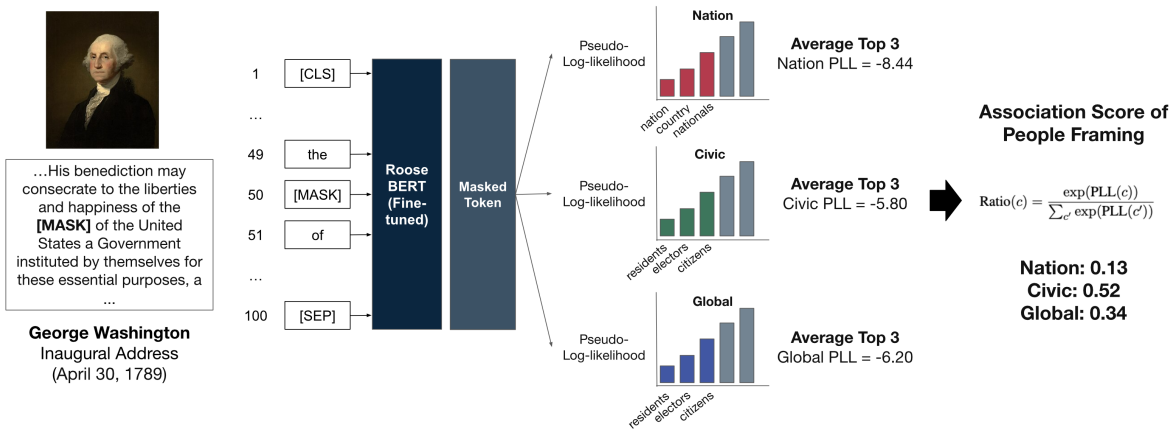

Figure 1: **Schematic illustration of the Masked Association Score computation.** For each occurrence of *people*, a 100-token context window is constructed and the target term is replaced with a [MASK] token, then encoded by RooseBERT (cased). For each of the three theory-driven frames—Nation, Civic, and Global—we compute a category-level pseudo-log-likelihood (PLL) by averaging the top-3 dictionary-word probabilities at the masked position, and convert the three PLL scores into competing association ratios via softmax normalization. The example is drawn from George Washington's inaugural address (April 30, 1789), for which the Civic frame receives the highest ratio (0.52).

sulting map exhibits coherent clustering by both historical era and individual president, suggesting that the embeddings encode rhetorical shifts across American political history in ways relevant to the downstream analysis (Appendix A.3).

2.3 Preprocessing

We included only speeches containing the token *people*[3], since this was the central object of analysis. As BERT-based models have a maximum input length of 512 tokens, processing full speeches directly is infeasible for many presidential addresses, which often far exceed this limit. Moreover, even within a single speech, the meaning of *people* can vary considerably with the local context in which it appears. To address both concerns, we constructed fixed-length windows of 100 tokens around each occurrence of *people*, with 50 tokens preceding and 50 tokens following the target word, providing the masked language model with sufficient local context to estimate fill-in probabilities. When fewer than 50 tokens preceded *people*, the window was adjusted to begin at the start of the encoded sequence. This preprocessing yielded a total of 29,424 contextual instances of *people* for analysis.

[3]Throughout this paper, "the people" refers to the conceptual object of analysis whereas *people* (in italics) denotes the token as it appears in the text.

3 Quantifying Masked Association Score

To analyze how "the people" is framed in presidential rhetoric, we introduce the Masked Association Score, a measure designed to capture associations with ideological categories that are implied by the surrounding context rather than stated explicitly. Card et al. (2022) employed a BERT-based masked language model, replacing mentions of *immigrants* with a [MASK] token to quantify the degree of dehumanization in metaphors for ethnic minority groups. Building on this approach, we examine how the rhetorical framing of "the people" has shifted over time in U.S. presidential speeches. Figure 1 provides an overview of the procedure.

3.1 Theory-Guided Dictionary Construction

Our typology of three frames builds on an established theoretical tradition in nationalism studies. Kohn (1944) was the first to systematize the key distinction between ethnic and civic forms of nationalism, later developed by Meinecke (1970) and Brubaker (1992). The *Nation* frame corresponds to what this tradition calls ethnic or cultural nationalism, in which peoplehood is defined through common heritage, ancestry, territorial origins, and ethno-cultural markers such as language, religion, and tradition (Kolstø, 2000; Janmaat, 2006; Reeskens and Hooghe, 2010; Lubbers and Coenders, 2017; Harrison and Bruter, 2011). The *Civic* frame instead defines membership through

shared institutional and legal norms, including constitutionally defined rights and duties and democratic participation (Brubaker, 1992; Jayet, 2012; Larin, 2020; Tamir, 2019). The *Global* frame extends membership beyond the state, drawing on cosmopolitan theories that posit moral obligations toward humanity as a whole and emphasize solidarity, cooperation, and universal brotherhood across sovereign boundaries (Kohn, 1932, 1934; Nussbaum, 1996; Appiah, 2018; de Sousa Santos, 2007; Tinsley, 2019; Mylonas and Tudor, 2021). Although the three frames may overlap in practice, they constitute analytically distinct poles against which we measure the degree of contestation in our data. To operationalize them, we identified category-specific seed words through an embedding-based procedure.

The candidate words for each category were derived through an embedding-based similarity search over RooseBERT-cased's vocabulary. For each category, a set of seed words was defined a priori on theoretical grounds: *Nation* (`nationals`, `state`, `nation`, `country`), *Civic* (`residents`, `taxpayers`, `voters`, `citizens`, `electors`), and *Global* (`peoples`, `mankind`, `humanity`, `humankind`, `inhabitants`). We then computed a category centroid by averaging the token embeddings of the seed words and L2-normalizing the result, using the static word embeddings from RooseBERT-cased's embedding layer. After computing cosine similarities between each vocabulary token and the three centroids, we inspected the top-50 tokens per category.

We constructed the final dictionary by retaining the seed words and augmenting each category with semantically coherent additions, while removing tokens more similar to a competing category than to their intended one. For instance, `inhabitants` was excluded from *Global* because it showed higher cosine similarity with *Civic* (0.663); conversely, `civilization` and `humans` were added to *Global*, and `nations` and `countries` to *Nation*. The resulting dictionaries are as follows.

- ***Nation***: `nation`, `nationals`, `country`, `nations`, `countries`
- ***Civic***: `citizens`, `voters`, `taxpayers`, `electors`, `residents`
- ***Global***: `peoples`, `mankind`, `humanity`, `humankind`, `civilization`, `humans`

Table 1: **Pairwise mean cosine similarities between category word embeddings.** Diagonal entries (in bold) reflect intra-category cohesion; off-diagonal entries report inter-category similarity.

	Nation	*Civic*	*Global*
Nation	**0.257**	0.187	0.212
Civic	0.187	**0.459**	0.380
Global	0.212	0.380	**0.577**

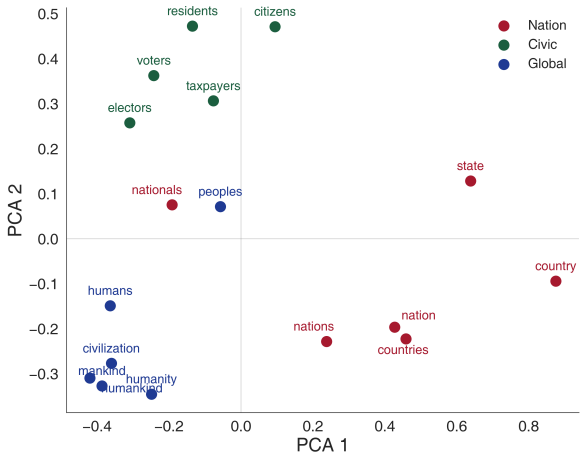

Figure 2: **PCA projection of the final dictionary word embeddings onto two dimensions**. Each point is a dictionary word, colored by category. The three categories occupy partially separable regions, with *Civic* and *Global* lying closer to each other than to *Nation*.

3.2 Dictionary Validation

To validate the resulting dictionary, we inspected the matrix of pairwise inter- and intra-category cosine similarities (Table 1). Intra-category similarities are consistently higher than inter-category similarities, with *Global* exhibiting the highest within-group cohesion (0.577) and *Nation* the lowest (0.257), reflecting the broader semantic range of the national-territorial concept. A PCA projection of the final word embeddings further confirms partial separability across the three categories in the two-dimensional subspace (Figure 2). Taken together, these results indicate that our theoretically motivated categories are not merely conceptual constructs but correspond to coherent, empirically separable regions of the embedding space, lending data-driven support to a dictionary grounded in the nationalism literature.

3.3 From Masked Predictions to Association Ratios

For every occurrence of *people* within a 100-token sliding window, we replaced the target word with a

[MASK] token and used RooseBERT to estimate the pseudo-log-likelihood (PLL) of each dictionary word filling the masked position (Salazar et al., 2020). For each masked instance, the category-level PLL was computed as the mean of the top-3 PLLs among the member words of category c:

$$\mathrm{PLL}(c) = \frac{1}{3}\sum_{k=1}^{3} \log P\left(w_k^{(c)} \mid w^{[\mathrm{MASK}]}\right) \quad (1)$$

where $w_k^{(c)}$ denotes the k-th highest-scoring dictionary word for category c, and $w^{[\mathrm{MASK}]}$ denotes the masked context. We then converted the three category-level PLLs into relative weights via softmax normalization:

$$\mathrm{Ratio}(c) = \frac{\exp(\mathrm{PLL}(c))}{\sum_{c'} \exp(\mathrm{PLL}(c'))} \quad (2)$$

which yields a continuous association ratio for each category at every instance. Intuitively, the three categories compete to fill the masked position, and the ratio captures how strongly each frame is favored relative to the others.

3.4 Robustness Checks

Our approach depends on three design choices: the number of probe words aggregated per category (k), the composition of the category dictionaries, and the choice of masked language model backbone. We examine the sensitivity of our measurements to each in turn, and find that the framing signal remains stable across all three.

Top-k sensitivity. We first assess the choice of $k = 3$ by comparing window-level correlations across $k \in \{1, 2, 3, 4, 5\}$. Scores computed with $k = 3$ and with $k \in \{4, 5\}$ are highly correlated across all three categories ($r > 0.95$), confirming that the results are stable across reasonable choices of k. By contrast, $k = 1$ yields substantially lower agreement, particularly for Global framing ($r = 0.54$), reflecting the rank instability of individual probe words across contexts. We therefore adopt $k = 3$ as a principled middle ground that balances representational coverage with robustness. The full correlation matrices are reported in Appendix A.1.

Dictionary sensitivity. We next verify that the results are not driven by any single word choice through a leave-one-out (LOO) analysis, in which each probe word is removed one at a time and the resulting change in category ratios (Δratio) is measured across all windows. In every case, removing a word lowers its own category ratio while the remaining categories absorb the shift, confirming that each word pulls in the expected direction. Although *citizens* and *peoples* carry relatively greater weight within the Civic and Global categories, no single removal reverses the overall framing pattern, indicating that the measurements capture a distributed semantic signal rather than hinging on any individual lexical choice. Full results are reported in Appendix A.2.

Cross-model robustness. Finally, we assess whether the framing signal is specific to RooseBERT's pretraining profile by replicating the scoring procedure with BERT-base-cased as an alternative backbone. Window-level rank agreement between the two models yields a mean Kendall's $\tau = 0.610$, with 52.9% of windows showing perfect rank correspondence across all three categories and only 2.0% showing complete rank reversal. These results confirm that the framing signal reflects a robust semantic pattern, one that generalizes across masked language model backbones rather than arising from idiosyncratic properties of a single pretrained model.

4 Results

4.1 Two Centuries of "the people" Framing

We first aggregated the instance-level scores to the speech level by computing the mean Nation, Civic, and Global ratios across all occurrences of *people* within each speech. These speech-level scores were then averaged by year to obtain annual framing estimates, to which we applied a 5-year centered rolling average to smooth out year-to-year fluctuations (Figure 3).

Figure 3 reveals a broad historical shift in how U.S. presidents have framed "the people" over time. We emphasize that our aim is a historically grounded interpretation of these patterns: the alignments we describe are correlational, and we make no claim that specific events drove presidential rhetoric or vice versa. Throughout the nineteenth century, Nation framing remained dominant while Civic framing gradually expanded as the polity absorbed new territories and diverse populations—reflecting the constitutional integration that bound an expanding republic, even as the franchise itself remained narrow (Smith, 2003). This civic expansion ultimately could not avert the constitutional crisis of the Civil War (1861–1865). The pattern reverses from the late nineteenth century onward:

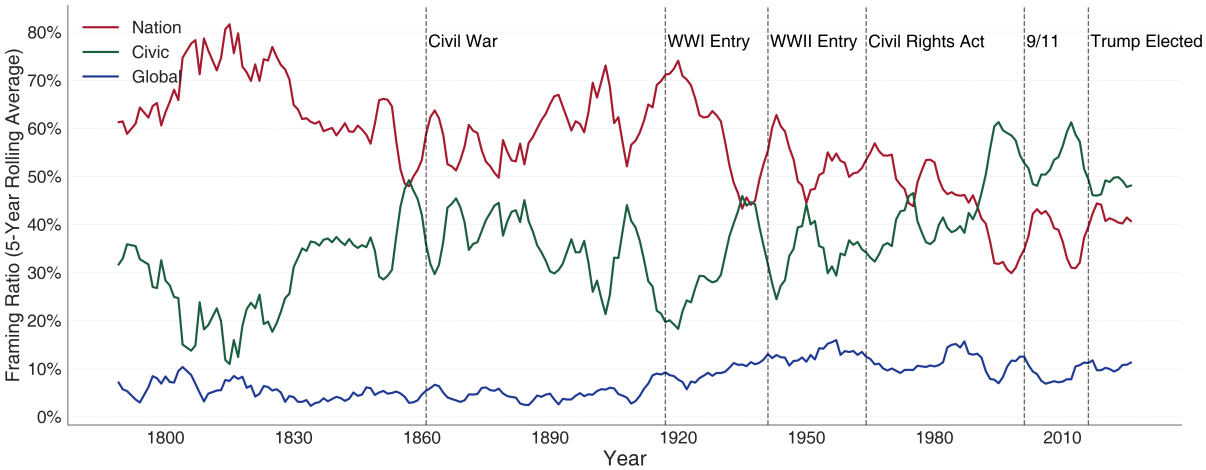

Figure 3: **Temporal evolution of Nation, Civic, and Global framing ratios in U.S. presidential speeches (1789–2026).** Solid lines denote 5-year centered rolling averages and vertical dashed lines mark major historical events. Nation framing dominates the nineteenth century before Civic framing overtakes it in the post-Cold War era, with a renewed nationalist turn after 2016.

as Reconstruction faltered, mass immigration provoked a nativist backlash, and imperial expansion and WWI mobilization mounted, Nation framing rose steadily while Civic framing receded, reaching a historical nadir of roughly 15% at the time of WWI entry (1917). "the people" was increasingly defined along ethnic and territorial lines rather than constitutional ones (Higham, 2002; Gerstle, 2017).

The mid-twentieth century marks a structural turning point. Following WWII, Civic framing rose steadily, converging with Nation framing around the Civil Rights Act era (1964) and eventually overtaking it in the post-Cold War period. By the 1990s, Civic framing had become the dominant frame, reflecting a broader shift toward inclusive, institutional conceptions of peoplehood in the aftermath of the Cold War. This dominance, however, was not uniform: Civic framing dropped noticeably after the September 11 attacks (2001), as Nation framing temporarily resurged amid national security mobilization. The trend reversed more sharply and durably after Trump's election (2016), with Nation framing rebounding and Civic framing declining—a renewed nationalist turn in presidential rhetoric that persisted beyond a single crisis moment (Bonikowski et al., 2021; Rowland, 2021).

Global framing remained comparatively low and stable throughout the period, fluctuating between 5% and 15%. This pattern is largely expected, since political elites, including U.S. presidents, address a primarily domestic audience. Even so, Global framing rose modestly between the post-WWII and Cold War periods, reflecting moments when presidents extended their rhetorical appeal beyond the national community.

4.2 Identifying Macro-level Predictors using Regression Analyses

To examine how the masked association scores covary with socio-economic and political indicators, we regress the annual Nation, Civic, and Global framing ratios on a broad set of variables using fractional logit regression, which is appropriate for proportional outcomes bounded between 0 and 1 (Papke and Wooldridge, 1996). Each year is treated as an independent cross-sectional observation, and we apply HAC (Heteroskedasticity and Autocorrelation Consistent) standard errors to account for potential residual autocorrelation across adjacent years. We stress that these estimates are associative rather than causal; they serve to corroborate that the RooseBERT-derived measures align with established macro-level indicators in theoretically expected ways.

We estimate two models for each framing. The *Long-term model* covers the full available period ($N = 231$ years) and includes the V-Dem Liberal Democracy Index together with a binary indicator for the pre-1960 period, which accounts for structural differences in speech availability and rhetorical context before the mid-twentieth century. The *Modern model* restricts the sample to post-1960 years with complete economic data ($N = 53$ years) and additionally includes log GDP per capita, trade openness (imports as % of GDP), and log net migration. We allow up to 15 lags in the long-term model

Table 2: **Fractional logit regression results, with HAC standard errors in parentheses.** The democracy index is the V-Dem Liberal Democracy Index; *Long* and *Modern* denote the long-term ($N = 231$) and modern ($N = 53$) models.

	Nation		Civic		Global	
	Long	Modern	Long	Modern	Long	Modern
Liberal Democracy	-1.822^{***}	-0.792	1.538^{**}	-0.136	2.047^{**}	5.448^{***}
	(0.586)	(0.949)	(0.664)	(1.071)	(0.995)	(1.391)
Pre-1960	-0.042		-0.023		0.487	
	(0.203)		(0.231)		(0.436)	
GDP (logged)		0.016		-0.109		0.576
		(0.467)		(0.491)		(0.482)
Trade Openness		-0.033		0.049^{*}		-0.105^{***}
		(0.025)		(0.027)		(0.035)
Net Migration (logged)		-0.265		0.328		-0.342^{*}
		(0.208)		(0.218)		(0.192)
Intercept	1.038^{**}	2.118	-1.010^{**}	-1.870	-4.593^{***}	-5.672^{***}
	(0.436)	(1.994)	(0.494)	(2.103)	(0.775)	(2.086)
N	231	53	231	53	231	53
Pseudo R^2	0.281	0.368	0.231	0.408	0.100	0.362

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

and up to 5 lags in the modern model, reflecting the different temporal spans of the two samples.

Table 2 presents the results. In the long-term models, the democracy index shows theoretically consistent associations across all three framings: higher democracy scores are associated with greater Civic framing (1.538^{**}) and lower Nation framing (-1.822^{***}), while Global framing shows a positive association (2.047^{**}). These patterns suggest that democratization corresponds to a broadening of peoplehood conceptions—strengthening civic-institutional identification and cosmopolitan orientation while weakening exclusive national identity.

In the modern models, the democracy index is no longer significantly associated with either Civic or Nation framing, but it retains a strong positive association with Global framing (5.448^{***}), suggesting that in the post-war period higher democratic consolidation corresponds to markedly more universalist rhetoric. Trade openness is positively associated with Civic framing (0.049^{*}) and negatively with Global framing (-0.105^{***}), indicating that economic openness does not translate straightforwardly into cosmopolitan orientation in presidential discourse. Net migration is negatively associated with Global framing (-0.342^{*}), while its associations with Civic and Nation framing do not reach statistical significance, suggesting that migration-related pressures may dampen universalist rhetoric rather than stimulate more inclusive appeals.

Figure 4 further illustrates the bivariate relationship between the Liberal Democracy Index and each framing ratio across all years. Civic framing exhibits a positive association with the democracy level (Pseudo $R^2 = 0.231$), whereas Nation framing shows a clear negative trend (Pseudo $R^2 = 0.281$), consistent with the long-term model results. Global framing shows a modest positive association (Pseudo $R^2 = 0.084$), reflecting its relative stability across the historical period, outside the democratic context captured in the modern model.

Taken together, these results validate the masked association scores against external indicators in theoretically expected directions, while illuminating how macro-level factors shape the framing of "the people" across different time scales. The long-term model surfaces century-scale patterns: democratization is robustly associated with a broadening of peoplehood, weakening exclusive national identification while strengthening civic-institutional and cosmopolitan conceptions. The modern model, in turn, brings the post-war dynamics into focus, showing how economic openness and migration flows shape the boundaries of "the people" in ways

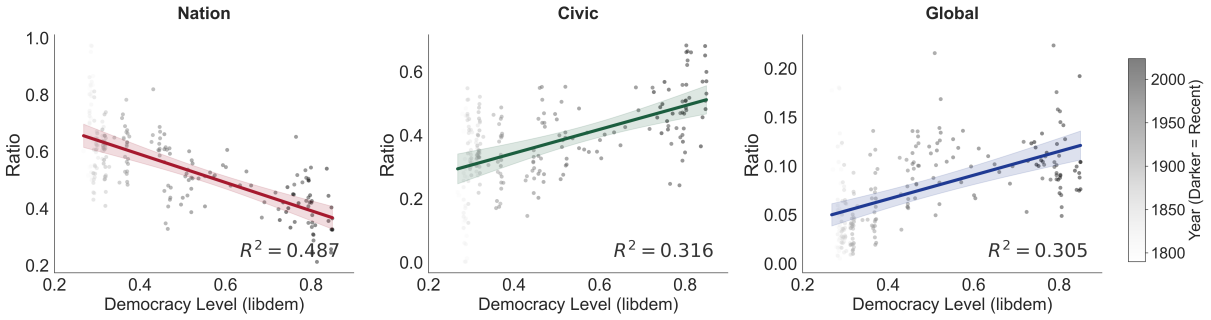

Figure 4: **Predicted framing ratios as a function of the Liberal Democracy Index, from the long-term model ($N = 231$).** Fitted lines are derived from fractional logit regressions, with shaded areas indicating 95% confidence intervals; each point is a single year, shaded by recency, with darker points denoting more recent observations. Civic framing rises with the democracy level while Nation framing declines, and Global framing increases modestly.

that complicate conventional accounts of globalization.

5 Discussion

Political elites frequently invoke "the people", yet its meaning remains contested and historically contingent. Treating political language as constitutive rather than merely descriptive, our analysis identifies systematic shifts in how "the people" is framed across three dimensions—Nation, Civic, and Global—and traces how these framings move in concert with broader political and economic developments.

Methodologically, the framework we introduce sidesteps several persistent limitations of prior approaches. By replacing each occurrence of a target term with a [MASK] token and reading off the relative likelihoods of theoretically guided probe words, the Masked Association Score captures local framing at the instance level without supervision, annotation, or generative inference. The resulting measurements are interpretable—anchored in transparent dictionaries grounded in political theory—and reproducible across model backbones, as our cross-model robustness checks confirm. This combination makes the framework well suited to contested concepts, whose meanings shift across contexts and historical periods in ways that conventional topic modeling or supervised approaches struggle to capture.

Substantively, our analysis surfaces a long-running contest among civic, ethnic, and cosmopolitan conceptions of "the people" in U.S. presidential rhetoric—one that shifts with episodes of war, immigration, civil rights, and populist resurgence, and tracks broader trajectories of democratic development and economic openness. Although our analysis cannot establish causal mechanisms, these alignments suggest that presidential rhetoric both reflects and constitutes changes in democratic practice. How political elites define "the people"—explicitly and implicitly—remains a central and evolving question in democratic politics, one that continues to shape the boundaries of political community and the distribution of rights and recognition.

More broadly, the framework generalizes beyond "the people". The same masked-language-model architecture, paired with theoretically guided dictionaries, can be applied to other foundational categories in political life—democracy, freedom, security—whose meanings are equally contested and historically variable. We hope this opens a path for the scalable, longitudinal study of how core political concepts are reshaped over time, complementing the qualitative scholarship that has long animated these debates.

Limitations

Several limitations of the present study warrant consideration. First, the masked association score relies on a fixed dictionary of category words, and the resulting scores are necessarily sensitive to the choice of these words. Although we validated the dictionary through embedding-based similarity search and cosine similarity analysis, the categories remain theoretically motivated constructs, and alternative word selections could yield different results.

Second, RooseBERT-cased was trained on political debate and speech corpora that extend beyond U.S. presidential speeches, which may introduce distributional biases into the model's probability estimates. Furthermore, as a static pre-trained model, it does not capture diachronic shifts in word mean-

ing; the semantic associations encoded in the model reflect a fixed snapshot of language use rather than the evolving meaning of words across centuries.

Third, the corpus is limited to textual records of presidential speeches, which are subject to selection biases in archiving and transcription, particularly for the nineteenth century. Earlier speeches are fewer in number and may overrepresent certain rhetorical occasions such as inaugural addresses, potentially skewing the temporal estimates for that period.

Finally, the regression analyses are associative rather than causal. While the co-variation between framing ratios and macro-political indicators is theoretically interpretable, confounding factors cannot be ruled out, and the direction of influence between presidential rhetoric and broader social conditions remains an open question.

Ethics Considerations

This study analyzes publicly available presidential speeches sourced from the American Presidency Project and the Miller Center. All texts are in the public domain, and no private or personally identifiable information is collected or processed.

The framing scores reported in this paper are descriptive and associative in nature. They are not intended to evaluate, rank, or judge the political legitimacy of any president, party, or administration, and should not be used for such purposes. We caution against causal or normative interpretations of the observed historical trends.

The automated nature of the measurement framework means that outputs reflect patterns in language model pretraining data as much as historical rhetoric itself. Downstream users of this framework should exercise caution when drawing conclusions about the intentions or ideologies of individual political actors based solely on model-derived scores. We recommend that quantitative outputs always be interpreted in conjunction with qualitative historical and political expertise.

The manuscript was prepared with the assistance of an AI assistant. All AI-generated content was reviewed and revised by the authors, who take full responsibility for the accuracy and integrity of the final work. All data, code, and model outputs will be made publicly available upon acceptance to ensure transparency and reproducibility.

References

Benedict Anderson. 2020. Imagined communities: Reflections on the origin and spread of nationalism. In *The New Social Theory Reader*, pages 282–288. Routledge.

Kwame Anthony Appiah. 2018. *The Lies That Bind: Rethinking Identity, Creed, Country, Color, Class, Culture*, 1 edition. Liveright Publishing Corporation, New York.

Vanessa B Beasley. 2011. *You, the people: American national identity in presidential rhetoric*, volume 10. Texas A&M University Press.

Bart Bonikowski, Yuval Feinstein, and Sean Bock. 2021. The partisan sorting of "america": How nationalist cleavages shaped the 2016 us presidential election. *American Journal of Sociology*, 127(2):492–561.

Rogers Brubaker. 1992. *Citizenship and Nationhood in France and Germany*. Harvard University Press, Cambridge, MA.

Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, 119(31):e2120510119.

Boaventura de Sousa Santos. 2007. Beyond abyssal thinking: From global lines to ecologies of knowledges. *Review (Fernand Braudel Center)*, 30(1):45–89.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Deborah Dore, Elena Cabrio, and Serena Villata. 2025. Roosebert: A new deal for political language modelling.

Gary Gerstle. 2017. *American crucible: Race and nation in the twentieth century*. Princeton University Press.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Sarah Harrison and Michael Bruter. 2011. *Mapping Extreme Right Ideology: An Empirical Geography of the European Extreme Right*, 1 edition. Palgrave Macmillan, London.

John Higham. 2002. *Strangers in the land: Patterns of American nativism, 1860-1925*. Rutgers University Press.

Jan Germen Janmaat. 2006. Popular conceptions of nationhood in old and new European member states: Partial support for the ethnic-civic framework. *Ethnic and Racial Studies*, 29(1):50–78.

Cyril Jayet. 2012. The ethnic-civic dichotomy and the explanation of national self-understanding. *European Journal of Sociology / Archives Européennes de Sociologie*, 53(1):65–95.

Hans Kohn. 1932. *Nationalism and Imperialism in the Hither East*. G. Routledge and Sons, Limited, London.

Hans Kohn. 1934. *Orient and Occident*. John Day Company, New York.

Hans Kohn. 1944. *The Idea of Nationalism: A Study in Its Origins and Background*. Macmillan, New York.

Pål Kolstø. 2000. *Political Construction Sites: Nation-Building in Russia and the Post-Soviet States*. Westview Press, Boulder, CO.

Stephen J. Larin. 2020. Is it really about values? Civic nationalism and migrant integration. *Journal of Ethnic and Migration Studies*, 46(1):127–141.

Marcel Lubbers and Marcel Coenders. 2017. Nationalistic attitudes and voting for the radical right in Europe. *European Union Politics*, 18(1):98–118.

Leland McInnes, John Healy, and James Melville. 2018. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Friedrich Meinecke. 1970. *Cosmopolitanism and the National State*. Princeton University Press, Princeton, NJ. Originally published as *Weltbürgertum und Nationalstaat*, 1907. Translated by Robert B. Kimber.

Harris Mylonas and Maya Tudor. 2021. Nationalism: What we know and what we still need to know. *Annual Review of Political Science*, 24:109–132.

Martha C. Nussbaum. 1996. Patriotism and cosmopolitanism. In Joshua Cohen, editor, *For Love of Country: Debating the Limits of Patriotism*, pages 2–17. Beacon Press, Boston.

Leslie E Papke and Jeffrey M Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics*, 11(6):619–632.

Tim Reeskens and Marc Hooghe. 2010. Beyond the civic–ethnic dichotomy: Investigating the structure of citizenship concepts across thirty-three countries. *Nations and Nationalism*, 16(4):579–597.

Robert C Rowland. 2021. *The rhetoric of Donald Trump: Nationalist populism and American democracy*. University Press of Kansas.

Alix Rule, Jean-Philippe Cointet, and Peter S Bearman. 2015. Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014. *Proceedings of the National Academy of Sciences*, 112(35):10837–10844.

Julian Salazar, Davis Liang, Toan Q Nguyen, and Katrin Kirchhoff. 2020. Masked language model scoring. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 2699–2712.

Vivien A Schmidt. 2008. Discursive institutionalism: The explanatory power of ideas and discourse. *Annual Review of Political Science*, 11(1):303–326.

Rogers M Smith. 2003. *Stories of peoplehood: The politics and morals of political membership*. Cambridge University Press.

Mary E Stuckey. 2023. *Defining Americans: The presidency and national identity*. University Press of Kansas.

Yael Tamir. 2019. *Why Nationalism*. Princeton University Press, Princeton, NJ.

Meghan Tinsley. 2019. Decolonizing the civic/ethnic binary. *Current Sociology*, 67(3):347–364.

John Zaller. 1992. *The nature and origins of mass opinion*. Cambridge university press.

A Appendix

A.1 Top-k Sensitivity Analysis

Figure 5 reports window-level Pearson correlations between framing scores computed under $k \in \{1, 2, 3, 4, 5\}$ for each category. Scores stabilize from $k = 3$ onward (highlighted with black borders), confirming that the top-3 averaging choice does not meaningfully affect the substantive results. Correlations between $k = 3$ and $k \in \{4, 5\}$ exceed $r = 0.95$ across all three categories, while $k = 1$ shows substantially lower agreement, particularly for Global framing ($r = 0.54$), reflecting rank instability of individual probe words across contexts.

A.2 Leave-One-Out Dictionary Sensitivity

Figure 6 presents the full leave-one-out (LOO) results. Each bar reports the mean change in category ratio (Δratio) when a given probe word is excluded from the dictionary, averaged across all windows. Words are grouped by their assigned category, separated by dashed lines.

For the Nation category, *nation*, *country*, *nations*, and *countries* produce the largest shifts ($|\Delta| \approx 0.05$–0.08), while *state* and *nationals* contribute more modestly. For the Civic category,

Figure 5: Window-level Pearson correlation matrices across top-k choices for Nation, Civic, and Global framing categories. Black borders indicate the $k = 3$ row and column selected for the main analysis.

citizens accounts for a disproportionately large shift ($\Delta \approx -0.16$), reflecting its high contextual fit in presidential rhetoric; the remaining words (*voters*, *residents*, *taxpayers*, *electors*) contribute more evenly. For the Global category, *peoples* carries the largest share, while *humans*, *humanity*, *mankind*, *civilization*, and *humankind* produce negligible shifts. Across all categories, the direction of change is consistent with theoretical expectations: excluding a word always reduces its own category ratio, with compensating increases absorbed by the remaining categories.

A.3 Speech Embedding and Topic Modeling

Figure 7 presents the UMAP projection of all speech windows containing *people*. Each window is encoded into a fixed-length vector via the CLS token of RooseBERT, and projected into two dimensions using UMAP (McInnes et al., 2018) with n_neighbors set to 15. Semantic clusters are identified through BERTopic (Grootendorst, 2022) using HDBSCAN with min_cluster_size set to 10, with cluster labels refined using GPT-4o-mini. The map reveals broad era-level separation, with Nation-oriented framing concentrated among earlier administrations and Civic framing more prominent in the modern era.

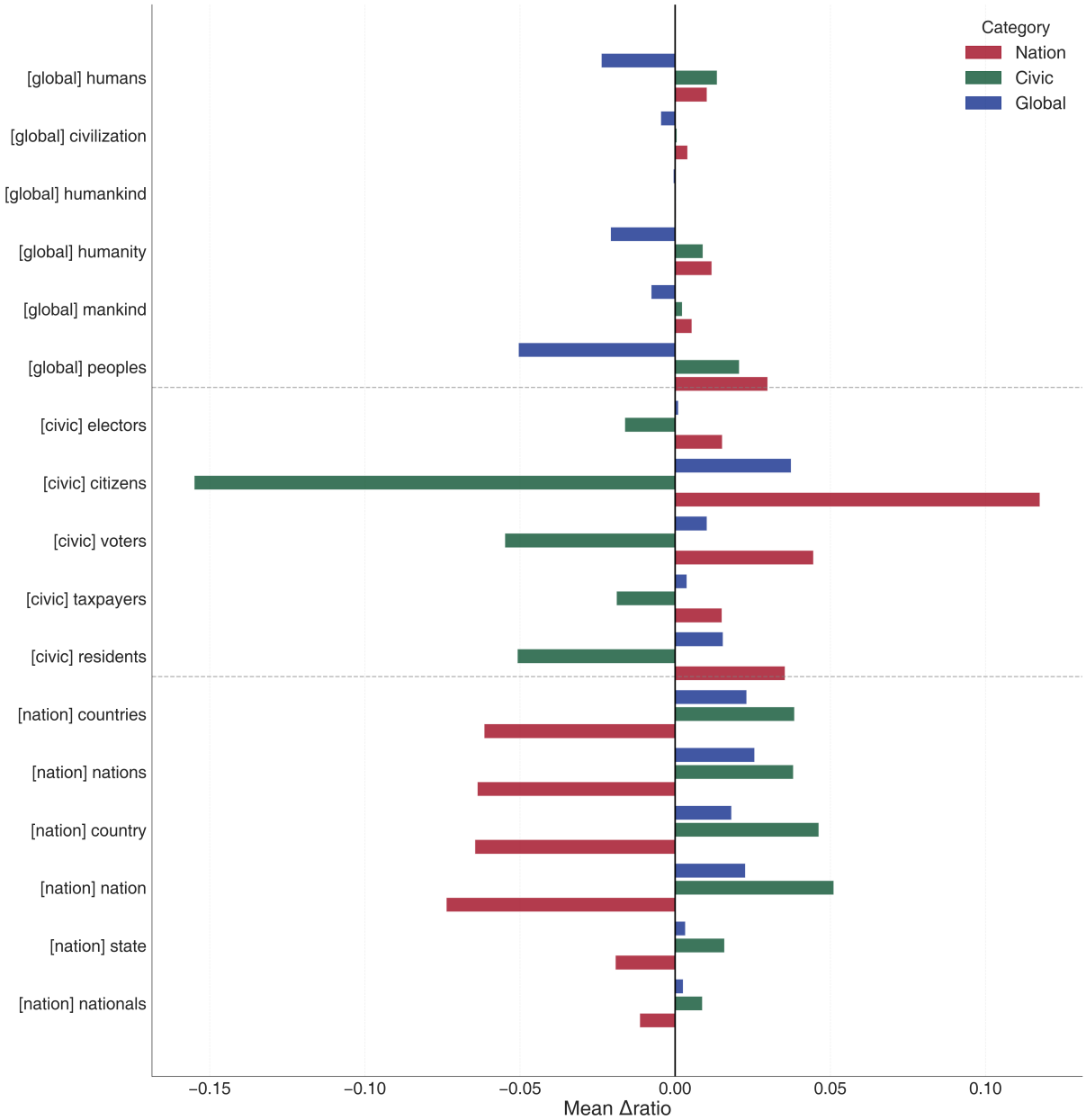

Figure 6: Leave-one-out sensitivity analysis. Each bar reports the mean Δratio when the target candidate word is excluded, averaged across all windows. Colors indicate the affected category. Words are grouped by assigned category (dashed lines). *Citizens* and *peoples* show the largest individual impact within the Civic and Global categories, yet omitting any single word shifts the magnitude but not the overall ranking of category ratios, indicating that the results do not hinge on any one lexical choice.

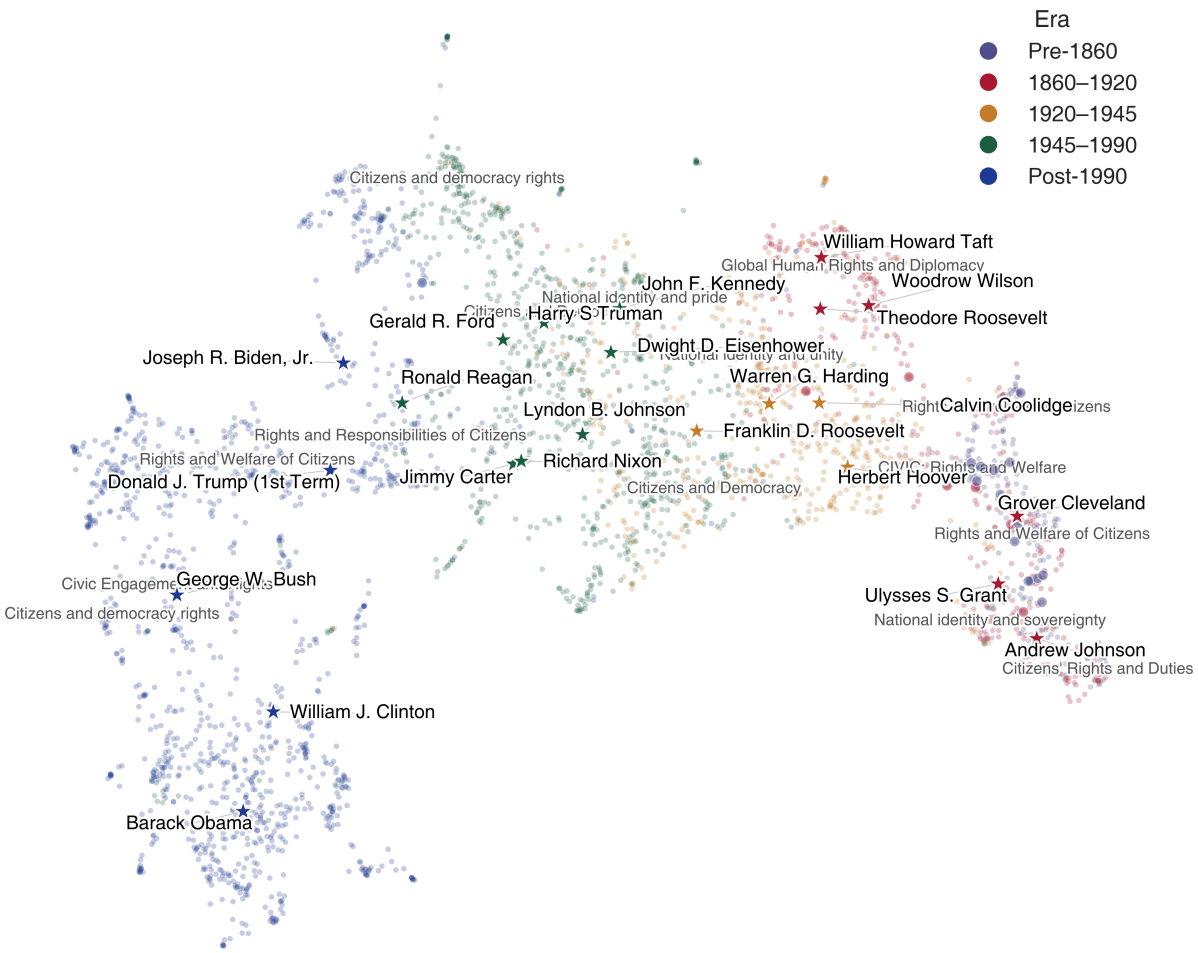

Figure 7: UMAP projection of RooseBERT [CLS] embeddings for all presidential speeches. Each dot represents a single speech, and stars mark the medoid (the most centrally located speech) of presidents with 30 or more speeches in the corpus. Colors indicate historical era. Italic labels denote the 15 largest BERTopic clusters, with labels refined using GPT-4o-mini. Speeches cluster coherently by both historical era and individual president, suggesting that the embeddings encode rhetorical shifts across American political history.