When Words Divide: A Computational View of Political Polarization

Roy Yitzchak

April 2026

Final project submitted in partial fulfilment of the requirements for the M.Sc. degree in the School of Computer Science of the Academic College of Tel Aviv-Yaffo

The research work for the final project has been carried out under the supervision of

Dr. Ella Rabinovich

THE ACADEMIC COLLEGE OF TEL AVIV-YAFFO

School of Computer Science - The Academic college of Tel Aviv-Yaffo, Israel

Abstract

Political polarization has become a defining feature of contemporary political discourse in the United States, raising important questions about how ideological divisions manifest and evolve in everyday language. This study investigates the dynamics of linguistic polarization by analyzing political discussions on Reddit over a fifteen-year period. Rather than relying on predefined lists of political terms, the analysis first identifies politically characteristic vocabulary through log-odds ratio analysis with an informative Dirichlet prior, contrasting political communities with non-political Reddit corpora.

To examine how political language differs between ideological groups, distributional semantic models based on Word2Vec were trained for eleven temporal slices of the data. Within each time period, tokens were marked by political affiliation, enabling direct comparison between Democratic and Republican contextual representations of the same word. Semantic divergence was quantified using a shift score defined as one minus the cosine similarity between the corresponding embeddings of each term. The resulting time series of shift scores was then analyzed using the Mann–Kendall test to detect monotonic trends in semantic divergence.

The results reveal a clear and systematic pattern of increasing linguistic polarization. Among words with sufficient temporal observations, the vast majority exhibit a statistically significant increase in divergence in contextual usage between the two political communities. Complementary qualitative analysis further highlights recurring patterns of asymmetric labeling, identity-based framing, and shifts in evaluative meaning across communities. Moreover, the temporal findings indicate that these semantic gaps are not static, but tend to expand over time. Taken together, these results provide converging quantitative and qualitative evidence that ideological polarization is reflected not only in political attitudes, but also in the evolving and increasingly divergent semantics of online political language.

Acknowledgments

I am sincerely grateful to **Dr. Ella Rabinovich** for supervising this work. From our very first interaction, her availability and support stood out, providing me with the necessary tools and knowledge to approach this field with confidence. I would like to thank her for being open to my suggestions and for her thoroughness in every detail of the research. Her guidance in maintaining a structured and documented methodology was crucial, not only for the completion of this project but for the benefit of future research in this area.

Content

1. Introduction

In recent years, **political polarization** has become one of the most widely discussed social and political dynamics shaping contemporary political life in the United States. In 2024, the *Merriam-Webster Dictionary* named "polarization" its word of the year, reflecting the growing public concern over deepening ideological divides. Broadly defined, polarization refers to the **increasing** ideological and emotional **distance** between individuals or groups, often manifested in opposing political beliefs, mutual distrust, and sharply differentiated social identities.

The rise of online communication, which became a major medium through which political polarization accelerated, has fundamentally transformed how political discourse unfolds. Since the early 2000s, social media platforms have become central arenas in which individuals discuss political events, share opinions, and engage in debate. Among these platforms, **Reddit** stands out as a particularly rich environment for studying political communication. Organized into topic-based communities known as **subreddits**, Reddit enables large groups of users to participate in ongoing discussions about political issues, often within ideologically aligned **communities**. These discussions provide a large-scale textual record of how political language is used and how it **evolves over time**.

Political polarization is commonly conceptualized in two distinct forms: ideological polarization and affective polarization. Ideological polarization refers to increasing divergence in policy preferences, political beliefs, and worldviews—for example, disagreements regarding immigration policy, healthcare systems, or gun regulation. Affective polarization, by contrast, concerns emotional hostility and distrust toward members of the opposing political camp. While prior computational work on political discourse typically examines affective signals or ideological aspects in isolation—often via sentiment analysis, emotion detection, or related semantic methods—this study focuses explicitly on ideological (thematic) polarization between groups, analyzing how it manifests in linguistic patterns, including shifts in semantics, contextual usage, and framing across communities.

Studying polarization computationally presents several methodological challenges. Most importantly, polarization is not a directly observable variable that can be easily labeled in text. Determining whether a specific post is "polarized" is inherently subjective and depends heavily on historical context, political events, and cultural interpretation. A straightforward supervised learning approach—training a classifier to identify polarized content or infer political orientation—would require large amounts of labeled data and may capture only superficial cues rather than deeper semantic differences. For this reason, unsupervised approaches that analyze patterns emerging directly from language usage offer a more flexible and generalizable framework for studying polarization in large textual corpora.

To address this challenge, this study adopted a linguistic modeling approach based on distributional semantics. The central intuition, famously articulated by the linguist J. R. Firth as *"You shall know a word by the company it keeps,"* suggests that a word's meaning can be inferred from the contexts in which it appears. **Word embedding models such as Word2Vec** operationalize this idea by representing words as dense vectors in a high-dimensional space, where semantic similarity corresponds to geometric proximity. Words that occur in **similar contexts** are therefore represented by **nearby vectors**.

Using this framework, we trained separate embedding models for political corpora associated with **liberal (Democratic)** and **conservative (Republican)** Reddit communities across multiple time periods. Rather than relying solely on a predefined set of politically obvious terms, we aimed to identify politically distinctive vocabulary in a **systematic and data-driven manner**. While certain words—such as *immigration*, *freedom*, *healthcare*, or *patriot*—serve as intuitive examples of politically relevant language, restricting the analysis to such manually selected terms would risk overlooking other important signals present in the data. To address this challenge, we first applied **log-odds ratio analysis with an informative Dirichlet prior**, a statistical method introduced by Burt L. Monroe and colleagues for identifying words that are significantly characteristic of one corpus relative to another.

This method enabled us to systematically identify vocabulary that is characteristic of political discourse as a whole, by contrasting the combined corpus of political communities (Democratic and Republican) with a collection of non-political Reddit communities. This step provided a principled and data-driven way to surface politically relevant terms without relying on manually constructed word lists, providing a systematic starting point for subsequent semantic analysis.

Once these politically characteristic terms were identified, we analyzed their contextual representations using Word2Vec embeddings trained separately for each time period. For each time slice, comments and posts from Democratic and Republican communities were combined into a single corpus while marking each token with a community-specific **suffix** (e.g., *patriot_**D*** and *patriot_**R***). This allowed the model to learn separate contextual representations for the same lexical item as used by each political group within the **same embedding space**.

For every politically relevant term, we computed a **shift score** defined as $1 - \cos(\mathbf{w}_D, \mathbf{w}_R)$, where the cosine similarity is calculated between the vector representations of the Democratic and Republican versions of the same word (e.g., *patriot_**D*** and *patriot_**R***). This procedure produced a sequence of shift scores across the eleven time periods, reflecting how differently the same term was used by the two ideological communities at each point in time. To determine whether these shifts exhibited a systematic temporal pattern, we applied the **Mann–Kendall test**, a non-parametric statistical method designed to detect monotonic trends in time-series data. This analysis allowed us to evaluate whether the semantic divergence between the two communities showed an *increasing trend*, a *decreasing trend*, or *no consistent trend* over time.

In addition to this quantitative framework, the study also incorporates a qualitative analysis of selected terms, examining how patterns of usage reflect broader forms of polarization, including asymmetric labeling, evaluative framing, and intergroup dynamics. Taken together, this framework provides a combined quantitative and qualitative approach for examining the evolution of linguistic polarization in online political discourse. By integrating statistical identification of politically characteristic vocabulary with diachronic embedding analysis and qualitative interpretation, the study evaluates not only whether the contextual meanings of political terms used by Democratic and Republican communities have become increasingly divergent over time, but also how these differences manifest in practice. The findings provide converging evidence that the semantic usage of many politically relevant terms has grown progressively more distinct between ideological groups, and that these differences tend to expand over time, contributing to a deeper understanding of how political polarization is reflected and reinforced through language.

All data and resources used in this study are publicly available at the following link:

https://drive.google.com/drive/folders/1h7grSeMm_hah9QDEZBSrkmongfCDFp3I?usp=sharing

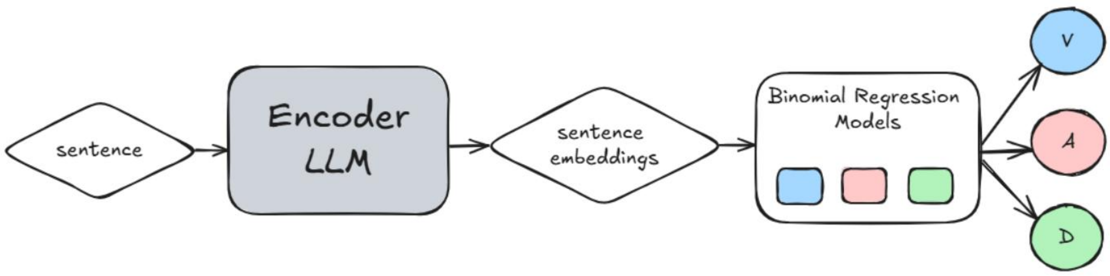
2. Related Work

Recent years have seen extensive research on political polarization, particularly in computational linguistics. Most prior work focuses on **ideological polarization**, defined as differences in viewpoints and language use between political groups. For example, Jensen et al. (2012) [2] showed that political language in the U.S. Congress has become increasingly polarized over time, while Demszky et al. (2019) [3] identified strong partisan linguistic differences in social media discourse.

Other approaches model polarization through semantic representations. Milbauer et al. (2021) [4] used word embeddings to capture divergence between online communities, and Sinno et al. (2022) [5] proposed a multidimensional ideological framework for analyzing political content. Additionally, Frimer et al. (2022) [6] demonstrated a rise in toxic and uncivil language in political communication, further indicating increasing polarization.

Beyond ideological differences, a complementary line of work examines **affective polarization**, which focuses on the emotional aspects of political discourse. Introduced by Iyengar, Sood, and Lelkes (2012) [7], this concept captures the tendency of individuals to express negative emotions toward opposing groups. Computational studies have operationalized this idea using sentiment and emotion analysis tools, such as the Valence–Arousal–Dominance (VAD) framework (Mehrabian and Russell, 1974) [8].

Goldin et al. (2025) [1] extend this direction by proposing a computational framework for measuring affective polarization in parliamentary discourse. Using a large corpus of Knesset proceedings, they combine sentence embeddings from a fine-tuned Hebrew encoder model with regression methods to predict VAD values. Their results show increasing affective polarization over time, as well as clear emotional differences between coalition and opposition speech.



In contrast, our work adopts a different approach to analyzing polarization. Rather than explicitly modeling emotional dimensions, we rely on distributional semantics to capture differences in language use across groups, focusing on how polarization—particularly ideological polarization—manifests in linguistic patterns.

Importantly, our work is complementary to affective approaches such as Goldin et al. (2025). While their method captures emotional dynamics, ours focuses on semantic and contextual distinctions in language use. Additionally, whereas Goldin et al. (2025) analyze formal, often heavily post-edited parliamentary speech from the Knesset, our study examines naturally occurring, unedited language from online social media discussions, capturing more direct and authentic forms of political expression. Together, these perspectives contribute to a more comprehensive understanding of political polarization.

3. Data

3.1 Data Source

The textual data used in this study were collected from Reddit, a large online discussion platform organized into thousands of user-created communities known as *subreddits*. Each subreddit focuses on a specific topic, interest, or ideological orientation, and discussions typically take place through posts and comments written by users.

For the purpose of analyzing ideological polarization in political language, the dataset was constructed from subreddits associated with major political affiliations in the United States. Left-leaning communities included subreddits such as **r/Democrats** and **r/Liberal**, while right-leaning communities included **r/Conservative** and **r/Republican**.

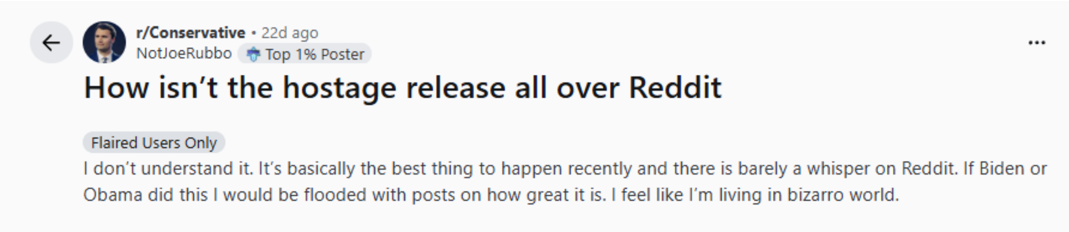
This design assumes that most users participating in these communities are broadly aligned with their respective political orientations. While some noise is expected—since users with opposing views may also contribute—active participation in ideologically oriented communities generally reflects a degree of affiliation with those communities.

Somewhat similar assumptions appear in prior work that infers user attributes from language and online behavior **(e.g., Shem-Tov and Rabinovich, 2025 [14]).** In their study, linguistic data is used to infer personality traits under the assumption that certain underlying characteristics are reflected—albeit imperfectly—in observed language. In a similar spirit, our approach assumes that political affiliation is partially reflected in community participation and language use. Although such assumptions introduce some noise, they provide a practical and commonly used proxy for analyzing group-level patterns in large-scale textual data.

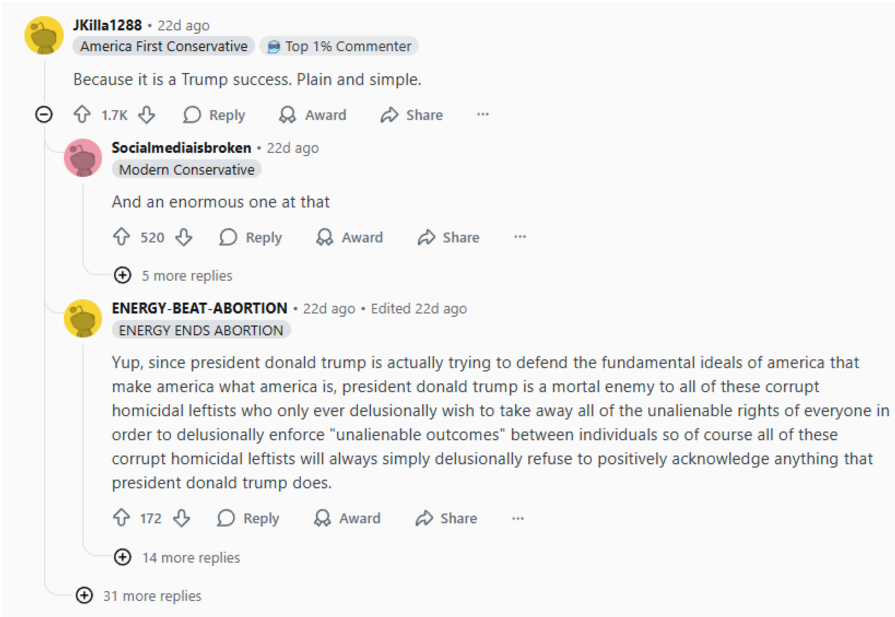
It is also important to note that political discourse on social media may be skewed toward more politically engaged or extreme users. **Preoțiuc-Pietro et al. (2017) [15]** show that political ideology is better understood as a nuanced spectrum, and that moderate users tend to exhibit weaker and less distinctive linguistic signals compared to more extreme individuals. As a result, large-scale social media datasets may overrepresent and oversimplified polarized voices. While this may limit coverage of the full ideological spectrum, it also makes linguistic signals of polarization more salient and easier to detect in the context of our research setup.

The raw corpus consisted of posts (submissions) and user comments collected over approximately fifteen years (2008–2023). The data were collected from Reddit using the Pushshift API (https://pushshift.io), when it was still publicly accessible, and stored in JSON format.

An example of a recent post:



Comments example to the given post above:



3.2 Data Preparation

The original data were stored in JSON format. For the purposes of the analysis, only the fields necessary for the study were retained, specifically the textual content and the creation date. The data were then converted into CSV format for easier processing.

As an initial filtering step, texts containing fewer than three words were removed, as such entries typically provide little semantic information. All remaining text was converted to lowercase to ensure consistent token matching across the corpus.

Posts and comments belonging to the same political community were merged into a single dataset per community. Sentence boundaries were identified using the *sent_tokenize* function from the NLTK library, which provides a pretrained tokenizer for detecting sentence boundaries.

3.3 Preprocessing

We experimented with several preprocessing configurations and ultimately chose to apply **minimal preprocessing** in order to preserve as much information as possible from the raw data. In particular, we did **not remove stop words and did not apply lemmatization**.

We also evaluated the use of **bigrams**, which requires identifying frequent word pairs (e.g., Donald_Trump) and tuning additional parameters that determine which word combinations should be merged into single tokens. However, incorporating bigrams introduced additional complexity without providing meaningful benefits for the analysis. Empirical evaluation in later stages of the project showed that models trained on corpora including bigrams exhibited slightly weaker performance on standard semantic evaluation benchmarks (measured using Pearson and Spearman correlations on the SimLex dataset). For this reason, we retained the corpus in its **unigram representation** for the final experiments.

3.4 Basic Statistics

Basic descriptive statistics were computed for the datasets, including counts of posts and comments, sentence counts, and length distributions. Detailed statistics are provided in Appendix A.

	Total posts (+comments)	Total sentences
Right wing	11,333,576	30,177,471
Left wing	1,141,642	3,381,569

Table 1: summary table with total posts(+comments) and total sentences for both wings

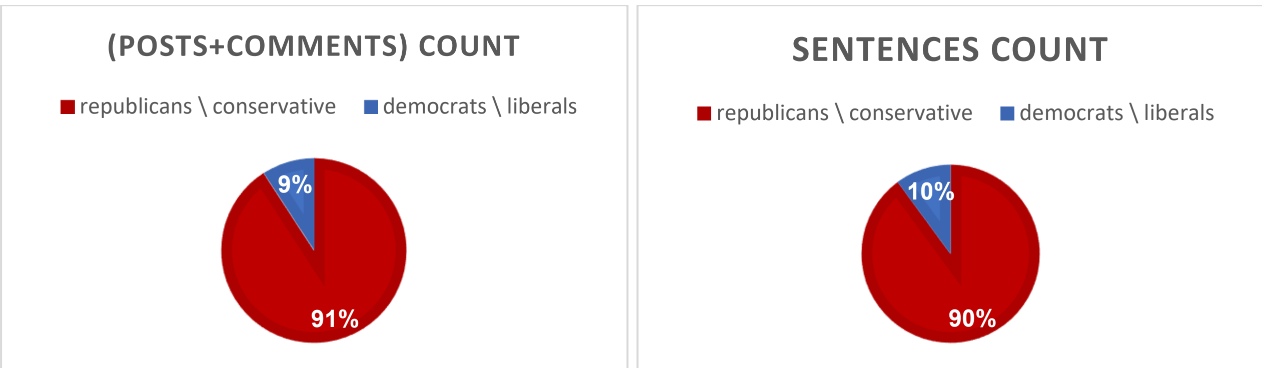
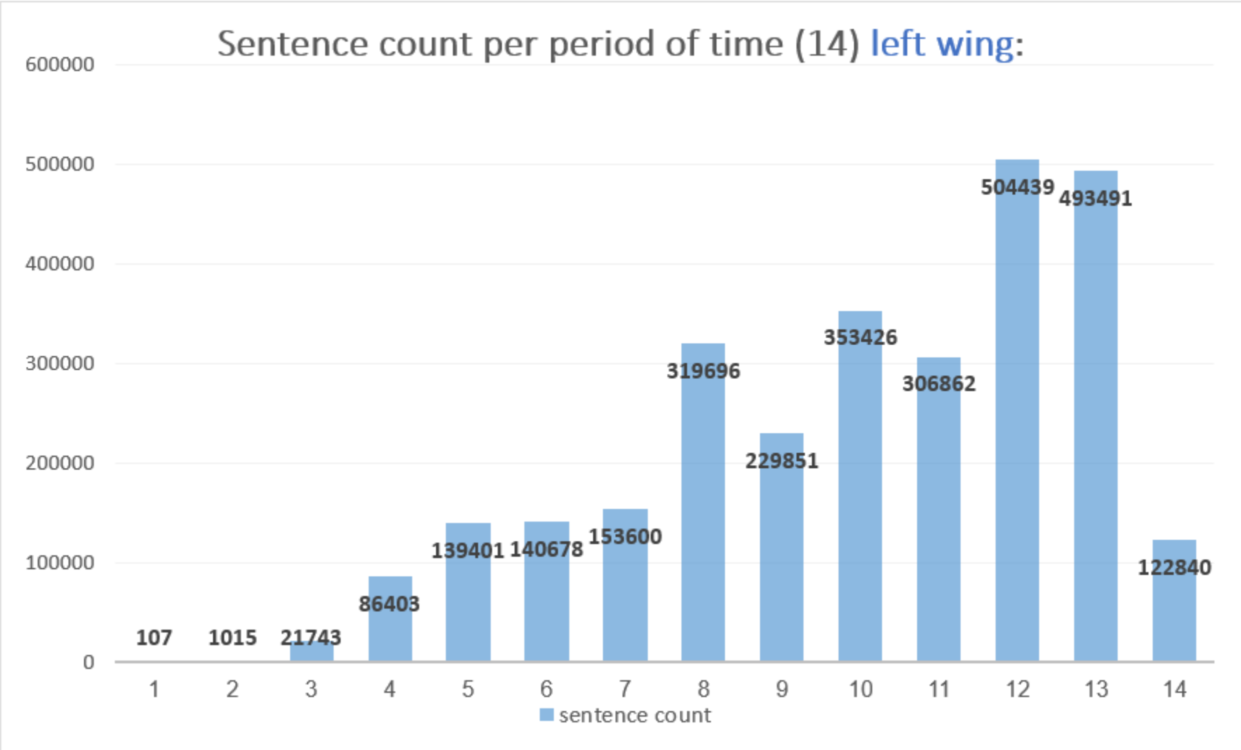
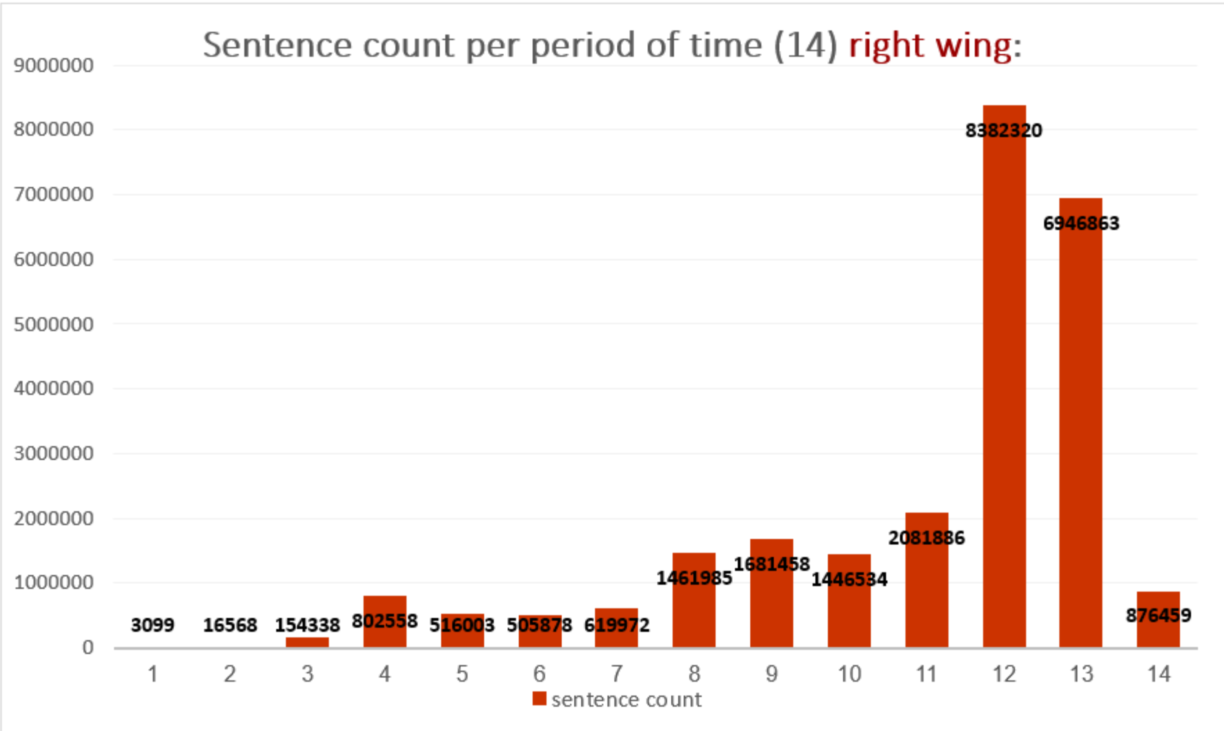


Figure1: summary pie-chart by community and wing





The temporal distribution of posts across the dataset is illustrated in **Figure 2 and 3**, which shows the yearly posting activity for each political community. We can see that, expectedly, with growing popularity of social media, the amount of data growths over the years.

Table 2: The above slices correspond to:

1	Before 2010	**8**	2016-2017
2	2010-2011	**9**	2017-2018
3	2011-2012	**10**	2018-2019
4	2012-2013	**11**	2019-2020
5	2013-2014	**12**	2020-2021
6	2014-2015	**13**	2021-2022
7	2015-2016	**14**	After 2022

4. Initial Semantic Modeling Approach

To analyze differences in language usage between ideological communities, we trained distributional word embedding models using **Word2Vec**, introduced by Tomas Mikolov et al [9][10]. Word2Vec is a **neural embedding** model that learns **vector** representations of words from their **surrounding context** in large corpora.

Word2Vec can be trained using two main architectures: Continuous Bag-of-Words **(CBOW)** and **Skip-Gram**. Both architectures learn word embeddings by predicting contextual relationships between words within a sliding window of neighboring tokens. Through this process, the model maps each word in the vocabulary to a dense numerical vector in a **high-dimensional space**.

Under this framework, words that appear in similar contexts tend to receive similar vector representations. As a result, the geometric relationships between vectors capture semantic and syntactic patterns in language. Previous work has demonstrated that these representations encode meaningful relationships that can be recovered through simple vector arithmetic—for example, the well-known analogy:

king − man + woman ≈ queen

These properties make word embeddings a powerful tool for analyzing semantic relationships in large text corpora, in fact, this property makes word embeddings particularly useful for analyzing ideological language differences. If two political communities use the same term in different contexts, the corresponding embeddings learned from their corpora will occupy different positions in the vector space, allowing such differences to be quantified through vector similarity measures.

In this study, we implemented and trained Word2Vec models using the Gensim library in Python. During training, the model first constructs a vocabulary from the corpus, after which embedding vectors are learned by predicting contextual relationships between words within a sliding window of neighboring tokens. The dimensionality of the embeddings and additional training parameters (such as context window size, minimum word frequency, and negative sampling settings) were specified as hyperparameters of the model. The selection and evaluation of these hyperparameters are discussed later in the next section.

After training the embedding models, the similarity between two words was measured using cosine similarity, which quantifies the angle between their embedding vectors. Cosine similarity is widely used in embedding-based semantic analysis because it captures the relative orientation of vectors independently of their magnitude.

Cosine similarity values close to 1 indicate that two words appear in very similar contexts, while values closer to 0 indicate little contextual similarity.

In practice, the similarity between two words can be computed directly using the implementation provided by the Gensim library.

Example:

```
word1 = "america"

word2 = "usa"

similarity_score = w2v_model.wv.similarity(word1, word2)
```

And viewing the nearest neighborhood (25 words) of a given word:

```
similar_words_25 = w2v_model.wv.most_similar(positive=["america"], topn=25)
```

These learned representations formed the basis for the subsequent analysis of ideological language differences and semantic divergence across political communities.

5. Identifying Politically Characteristic Vocabulary

5.1 Initial Attempts at Identifying Politically Associated Vocabulary

5.1.1 Manual Drafting a List of Political Associated Keywords

At the early stages of the project, we manually constructed a preliminary **list** of words seemingly associated with political discourse, such as patriot, freedom, immigration, drugs, peace, war, Trump and similar politically loaded expressions.

We hypothesized that our list does not fully capture the breadth of terms relevant to U.S. political discourse, as many important words did not come to mind during its construction.

5.1.2 Attempted Cross-Space Alignment that **was Discarded**

To systematically identify words whose meanings diverged most strongly between political communities, we also experimented with aligning the embedding spaces produced by the two models.

Each Word2Vec model generates its own vector space, meaning that the embeddings learned from the left-wing corpus and those learned from the right-wing corpus cannot be directly compared. To address this, we applied **Orthogonal Procrustes Alignment** by Schönemann et al. [11], a method that finds a rotation matrix that approximately align one embedding space with another while preserving vector distances and angles.

After alignment, vectors corresponding to the same lexical item in the two corpora could be directly compared using cosine similarity. We computed cosine similarities for all shared words and examined the terms with the **largest** semantic distances.

While this method did produce words with large distances between embeddings, the results were dominated by **non-informative** or irrelevant tokens (for example wikipedia, ykonw, and various proper names). Identifying meaningful political terms from this long list required extensive manual filtering and ad-hoc decisions. As a result, this approach did not provide a principled method for identifying politically meaningful vocabulary, and it was therefore **discarded**.

5.2 Identifying Politically Characteristic Vocabulary – Statistical Approach

In the last two sub-section we concluded that:

A suitable systematic statistical method was required to identify politically distinctive vocabulary in a principled way, rather than relying on manually selected word lists.

To address this, we adopted a statistical approach designed specifically for comparing word usage across corpora.

5.2.1 Log-Odds Ratio with an Informative Dirichlet Prior

We adopted **log-odds ratio analysis with an informative Dirichlet prior**, introduced by Monroe, Colaresi, and Quinn (2008) in *Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict* [13].

This method is widely used in NLP research for identifying words that are characteristic of one dataset compared to another (e.g., [16],[17],[18]).

A straightforward approach for comparing corpora is to examine raw word frequencies. However, such comparisons suffer from several limitations: they are sensitive to corpus size, tend to overemphasize common words, and may assign high importance to rare words that appear only a few times.

Another widely used measure is TF–IDF, which captures how important a term is within a document while discounting terms that are common across documents. While TF–IDF improves upon raw frequency counts, it does not directly measure whether a word is **statistically more associated with one corpus than another**.

Log-odds ratio analysis addresses this limitation by directly comparing relative word usage between corpora while incorporating a smoothing prior that stabilizes estimates for low-frequency terms. The informative Dirichlet prior introduces pseudo-counts that reflect overall word usage, preventing rare words from appearing artificially significant.

The method therefore identifies **characteristic words**—terms that are statistically distinctive between corpora rather than merely frequent. A positive score indicates that a word is more strongly associated with one corpus, while a negative score indicates stronger association with the other.

Importantly, raw effect sizes alone can be misleading: a difference based on thousands of observations is far more reliable than the same difference based on only a few occurrences. The log-odds framework addresses this by comparing **relative word usage between corpora, conditioned on how common the word is overall**, rather than relying on raw counts. Conceptually, the method computes the difference between the log-odds of a word in each corpus (incorporating Dirichlet prior), and normalizes this difference by its estimated variance. This variance reflects an intuitive property: rare words have higher uncertainty, while frequent words have more stable estimates. The resulting statistic is therefore a **z-score**, representing how strongly a word is skewed toward one corpus relative to the other, after accounting for corpus size and low-frequency noise.

Because these z-scores are approximately standard normal under mild assumptions, they are directly comparable across words and datasets. This makes it possible to rank words by their scores and identify those that are most strongly and reliably **characteristic** of a given corpus.

5.2.2 Extracting Political Vocabulary Using a Neutral Reddit Corpus

To obtain a unified set of politically characteristic words, we instead compared political discourse against a collection of **non-political Reddit communities**.

In this setting, the left-wing and right-wing corpora were combined into a single corpus representing political discourse. This combined corpus was then contrasted with a separate dataset constructed from general-purpose Reddit communities, including:

- *LifeProTips*
- *Showerthoughts*
- *relationship_advice*
- *socialskills*
- *technology*

These datasets were merged into a single corpus representing general, non-political language.

Applying the log-odds analysis in this configuration produced a ranked list of words with associated z-scores. Words with high positive scores were strongly characteristic of political discourse, while words with large negative scores were characteristic of general Reddit language.

The results aligned well with expectations. At the top of the ranked list, we observed clearly political terms such as **trump, biden, vote, democrats**, and **government**, while at the opposite end we observed general-purpose words such as **relationship, friend**, and **advice**.

side	log_odds_z	word
leftAndright	351.0858104	the
leftAndright	318.5771887	trump
leftAndright	211.5509191	they
leftAndright	196.1528724	biden
leftAndright	192.1044387	itâ€™s
leftAndright	182.5067517	vote
leftAndright	179.1422561	democrats
leftAndright	176.6627044	government
leftAndright	171.9181104	conservative

side	log_odds_z	word
TheMixDataSet	-223.1111143	me
TheMixDataSet	-240.5582052	you're
TheMixDataSet	-245.1298672	friends
TheMixDataSet	-280.1085429	she
TheMixDataSet	-284.1089816	relationship
TheMixDataSet	-300.6789393	my
TheMixDataSet	-364.6751795	her
TheMixDataSet	-380.9509875	your
TheMixDataSet	-396.4833391	i

Tables 3 and 4: Top 10 words from each side with their log odd z score – the higher the absolute score, the more associated is a word with the dataset. Also, We can notice that further filtering is necessary.

5.3 Filtering and Final Vocabulary Construction

From the ranked output, we focused on the top portion of the list **(from the political side),** (approximately 2,500 words) and performed a filtering process to construct a clean and meaningful vocabulary set.

This process initially involved reviewing the words through automated and manual steps and removing clearly irrelevant tokens, such as formatting artifacts, malformed text (e.g., encoding issues), and non-informative function words. Duplicate forms (e.g., punctuation variants such as *patriot* and *patriot.*) were also unified.

After this stage, the vocabulary was reduced to 1,017 candidate terms.

A second filtering step was then applied to ensure statistical reliability. Words that did not appear sufficiently often in both political corpora were removed. In particular, words with fewer than 100 occurrences in either corpus were excluded.

Following this filtering, the final vocabulary consisted of **851 politically relevant terms**, which **form the basis for all subsequent analysis in the study**.

A partial list of these terms is provided in **Appendix C**, as they constitute a central component of the methodological pipeline.

6. Word2Vec Model Evaluation

The Word2Vec training process involves several **hyperparameters**—including embedding dimensionality (=**vector size**), context **window size**, minimum word frequency, subsampling thresholds, and negative sampling parameters—which influence the quality of the learned embeddings. A detailed description of these parameters and the evaluated configurations is provided in **Appendix B**.

To strengthen and evaluate the semantic quality of the Word2Vec models, we tuned their hyperparameters using the **SimLex-999 benchmark dataset [12]** and explicitly validated their ability to capture genuine **semantic similarity**, rather than mere topical relatedness or association.

SimLex-999 is a human-annotated dataset designed specifically to measure true semantic similarity between word pairs, rather than simple association. Each pair of words in the dataset is assigned a similarity score by human annotators, reflecting how semantically similar the words are. For example, pairs such as car–automobile receive high similarity scores, while pairs like car–road, despite being related, receive lower scores.

To find the best combination of hyperparameters for training word embeddings from our data, cosine similarity scores generated by the embeddings were compared against human similarity judgments from SimLex-999. The agreement between model outputs and human annotations was quantified using **Pearson** and **Spearman rank correlation** coefficients. While Pearson measures the strength of the linear relationship between the two datasets, Spearman evaluates the consistency of their ranking order.

To identify suitable training configurations, we performed a **grid search** over multiple Word2Vec hyperparameter settings as described above. Grid search systematically evaluates combinations of model parameters, in order to identify **configurations** that produce the best semantic similarity performance by our evaluation strategy.

Among the tested configurations, the **best results** achieved approximately **0.42** Pearson correlation and **≈0.40** Spearman correlation on the SimLex-999 benchmark. Considering the moderate size and domain-specific nature of the Reddit corpus, these scores were deemed satisfactory and indicated that the embedding models captured meaningful semantic relationships. The full set of evaluated configurations and their corresponding Pearson correlation scores is provided in **Appendix B**. The best-performing configuration was: vector size = 300, window size = 2, subsampling = 6e-05, negative sampling = 20, and minimum word frequency = 100.

7. Measuring Ideological Semantic Shift (Preliminary Analysis)

Having identified a set of 851 politically characteristic terms, the next step was to measure how their semantic usage differs between ideological communities.

To enable a direct comparison within a shared embedding space, we introduced a **simple transformation**. Each occurrence of a word in the Democratic corpus was marked with the **suffix _D**, while occurrences in the Republican corpus were marked with **_R**. For example, the word *patriot* was represented as **patriot_D** and **patriot_R** depending on the source corpus.

The two corpora were then combined into a single dataset, and a Word2Vec model was trained using the previously selected hyperparameter configuration. Under this construction, the model treats patriot_D and patriot_R as distinct tokens, while still learning their representations within the same embedding space. This enables a direct comparison between the contextual meanings of the same lexical item as used by different political communities.

For each word in the filtered vocabulary, we extracted the corresponding embedding vectors for its Democratic and Republican variants and computed their cosine similarity. The **semantic shift** was then defined as:

$$shift(w) = 1 - cos(\overrightarrow{W_D}, \overrightarrow{W_R})$$

A small shift indicates similar contextual usage across communities, while a larger shift reflects a greater divergence in meaning or association.

The analysis was performed over all 851 words, and the results were stored in a structured dataset (**semantic_shift_scores.csv**), containing the word, its cosine similarity, and the corresponding shift score. The words were then sorted according to their shift values, allowing identification of terms exhibiting the largest semantic divergence between ideological groups.

In addition, word frequency counts in each corpus were retained (into the CSV data) to allow potential filtering based on occurrence thresholds. While this information was not used as a primary filtering criterion in this stage, it provided a safeguard against over-interpreting high shift values arising from extremely low-frequency words.

The results revealed a set of words with substantial semantic divergence between political communities. For example, terms such as *corona*, *lefty*, *neocons*, and *marxist* exhibited relatively large shift values, indicating differences in contextual usage across the two corpora.

word	cosine_similarity_R_vs_D	semantic_shift
corona	0.5121911	0.48780888
discord	0.5125213	0.48747867
lefty	0.53433526	0.46566474
appears	0.5546146	0.4453854
neocons	0.55752826	0.44247174
lefties	0.55958426	0.44041574
maxine	0.5674614	0.43253863
senile	0.5681868	0.43181318
radicals	0.5755013	0.42449868
vaccinations	0.5786226	0.42137742

Table 5: top 10 words by semantic shift (see Appendix D for extended list)

This initial analysis was performed on the full dataset without distinguishing between time periods. While exploratory in nature, it provided an important first indication that semantic differences between ideological communities are detectable using the proposed framework. Moreover, it allowed us to identify candidate words that may play a central role in subsequent temporal analysis.

8. Diachronic Semantic Shift and Trend Analysis

8.1 Temporal Data Construction

To examine semantic change over time, the dataset was partitioned into 11 time periods (rather than the initial 14) spanning 2008–2023. This configuration was chosen to balance temporal resolution with the need for sufficiently large corpora in each period for reliable Word2Vec training. In particular, earlier data was aggregated to ensure adequate coverage, as very small corpora do not provide stable contextual representations.

8.2 Semantic Shift Computation Over Time

For each of the 11 time periods, we applied the same procedure described in the previous section. At each time point, separate Democratic and Republican texts were marked using the _D and _R suffixes, merged into a single corpus, and used to train a Word2Vec model.

For each word in the final vocabulary (851 terms), we computed the semantic shift between its Democratic and Republican representations within each time period:

$$shift(w) = 1 - cos(\overrightarrow{W}_D, \overrightarrow{W}_R)$$

This resulted in a **time series of shift values** for each word:

$$[shift_1(w), shift_2(w), ..., shift_{11}(w)]$$

These values were stored in the file **semantic_shift_over_time_ORDERED.csv**.

However, not all words appeared in all time periods. In particular, words associated with specific historical events (e.g., **corona**, **pandemic**) only emerged in later years. Missing values were therefore recorded as NaN. Importantly, such words are not necessarily uninformative; rather, their absence reflects genuine temporal emergence.

To quantify this effect, we computed the distribution of missing values across words.

Number of NaNs	Number of words
0	570
1	73
2	47
3	33
4	30
5	24
6	18
7	18
8	15
9	14
10	8
11	0
Total	**850**

Table 6: counting words per NaN, we can see that most of the words appear at all periods – they have 0 NaN

8.3 Mann–Kendall Trend Test

To determine whether semantic shifts exhibit a consistent trend over time, we applied the **Mann–Kendall Test [19][20]**, a non-parametric method commonly used for detecting monotonic trends in time-series data.

The Mann–Kendall test evaluates the null hypothesis that a sequence has no monotonic trend, against the alternative hypothesis that it exhibits either an increasing or decreasing trend. The test does not assume normality and is robust to noise, making it well-suited for analyzing semantic shift sequences.

Because the test requires a sufficient number of observations, it was applied only to words with at least **9 valid** time points. Words with fewer observations were labeled as **insufficient data**. Although this corresponds to a relatively small number of temporal observations (9–11), prior work indicates that the Mann–Kendall test can be applied to short time series, with recommended minimum lengths of approximately 8–10 observations for reliable trend detection (see Mann–Kendall trend test overview).

The test produces several outputs. The most relevant for this study are:

- **z-score (Z):** measures the strength of the trend relative to noise
- **p-value (p):** indicates statistical significance **(we used a standard threshold of $p < 0.05$)**

- **trend:** categorical result (*increasing*, *decreasing*, *no trend*)
- **Sen's slope:** estimates the magnitude of change over time

The z-score serves as the primary indicator of trend strength, while the slope provides information about the rate of change.

8.4 Trend Analysis Results

Applying the Mann–Kendall test to all words yielded the following distribution:

- Increasing trend: **119 words (14.00%)**
- Decreasing trend: **5 words (0.59%)**
- No trend: **566 words (66.59%)**
- Insufficient data: **160 words (18.82%)**

As expected, not all examined political terms (N = 851) exhibit statistically significant trends, since uniform change across all terms is unlikely. However, a substantial proportion (14%) shows increasing semantic divergence over time.

Among the **124 words with a statistically significant trend**, **119 (≈96%) exhibited an increasing trend**.

This asymmetry suggests a consistent directional pattern, whereby semantic divergence between ideological communities tends to increase over time, providing evidence for a linguistic dimension of political polarization.

Word:	secular	ideological	invasion	abiding	arabia	bigot	victims	soldiers	genocide	shootings
Z-score	3.219	3.041	2.958	2.862	2.814	2.802	2.646	2.646	2.646	2.491
p-val	0.0012	0.0023	0.0030	0.0042	0.0048	0.0050	0.0081	0.0081	0.0081	0.0127

Table 7: Examples of words with strong increasing trends

8.5 Two-Era Comparative Analysis

In addition to the fine-grained temporal analysis, we performed a complementary coarse-grained comparison by dividing the dataset into two broader time periods:

- **Early period:** 2009–2018
- **Late period:** 2018–2023

Word embeddings were trained separately for each period, and semantic shifts were computed using the same methodology. The difference between the two periods was then calculated:

$$\Delta(w) = shift_{late}(w) - shift_{early}(w)$$

This coarse-grained analysis complements the findings of the Mann–Kendall test, serving as an additional tool, providing further support for the observed increase in semantic polarization over time, and being used to identify additional high-shift words for the qualitative analysis that follows.

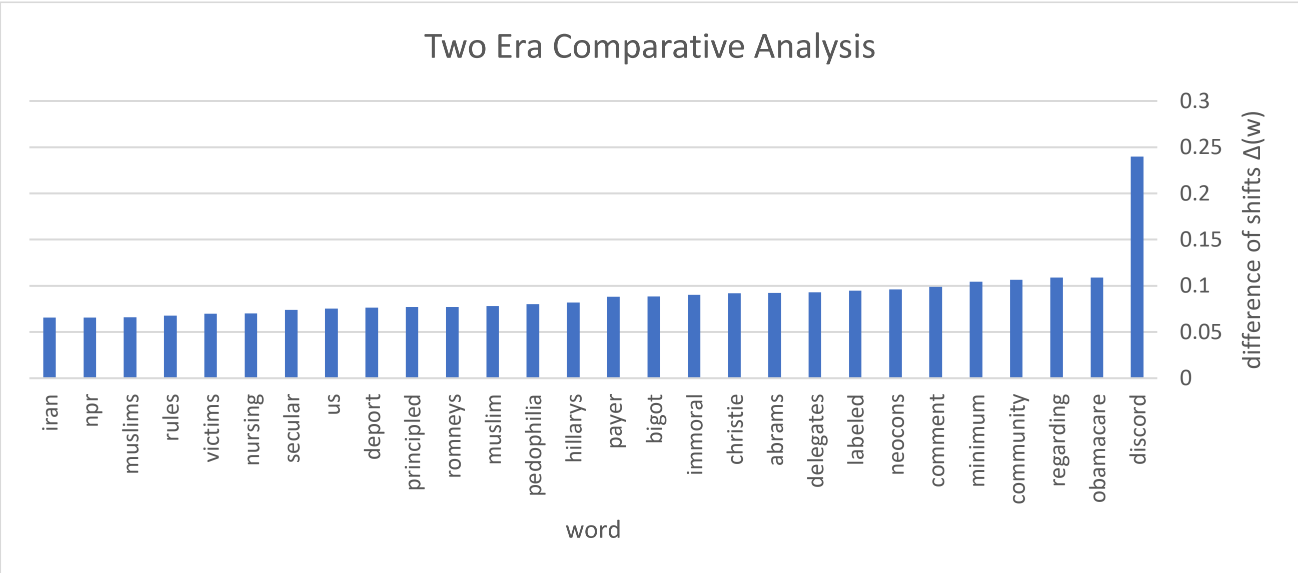


Figure 4: the top 28 words which has the largest difference in the two era shifts. The entire data can be accessed through the file "**semantic_shift_two_eras_sorted.csv**"

9. Qualitative Analysis of Semantic Divergence and Final Conclusion

Following the establishment of the methodological framework, data preprocessing, model evaluation, and the application of statistical techniques such as log-odds analysis and the Mann–Kendall trend test, we have accumulated a substantial body of quantitative evidence. Through extensive exploratory analysis, we extracted and curated a set of politically relevant terms from the corpora, which were subsequently filtered and organized for further examination.

For each selected term, we computed measures of semantic shift, intended to capture the degree of semantic distance between communities, as well as changes over time. These computations were further structured temporally, allowing us to distinguish between early and later periods. Collectively, these steps constitute a **quantitative analysis**, from which initial conclusions can already be drawn. In particular, the results indicate clear trends consistent with the research hypothesis, suggesting increasing semantic divergence between ideological communities.

However, quantitative results alone do not fully establish the nature or quality of this divergence. The present section therefore introduces a **complementary qualitative analysis.** This analysis serves a descriptive and interpretive role, aiming to examine whether the observed statistical patterns correspond to meaningful semantic differences. Specifically, we seek to determine whether the identified "semantic shifts" indeed reflect changes in ideological or conceptual meaning, or whether they may instead arise from artifacts, noise, or superficial contextual variations.

Given the scale and complexity of the data, a fully exhaustive qualitative inspection is beyond the scope of this study, and a focused sample is sufficient for the present analysis. The dataset includes a large number of terms, many of which are deeply embedded in the specific context of U.S. political discourse, and may require domain expertise to interpret accurately. In addition, qualitative interpretation inherently involves a degree of subjectivity. For these reasons, we focus on a subset of representative and interpretable cases—namely, terms that exhibit clear, prominent, and meaningful patterns of divergence. These examples are selected primarily from among the strongest quantitative results.

For each term, a large number of neighboring words can be identified in the embedding space. Not all of these neighbors are equally relevant for interpretation. Accordingly, in the examples presented below, we include only a subset of neighbors that best illustrate the observed patterns. These representations are therefore partial, but they are chosen to faithfully reflect the broader conclusions. It is important to note that the resulting interpretations are approximate and illustrative in nature, rather than precise or exhaustive, as is typical in qualitative analysis.

The qualitative analysis is based on data generated in earlier stages of the pipeline. Specifically, we utilize outputs from the following files: **semantic_shift_scores.csv**, **semantic_shift_two_eras_sorted.csv**, and **mann_kendall_semantic_shift_SORTED.csv**. From these sources, we selected several dozen candidate terms for inspection. For each term, we examined its nearest neighbors across different models and conditions, and ultimately retained a smaller set of particularly clear and informative examples.

The analysis is conducted in two complementary parts. **First**, we examine semantic divergence across communities, focusing on terms with high divergence scores. **Second**, we analyze **temporal** semantic shifts, considering both statistically significant trends and substantial differences between early and late representations **(Era A, Era B).** For this purpose, neighbors were extracted from four models: left (early), left (late), right (early), and right (late). This setup allows us to observe how meanings vary both across communities and over time.

9.1 Cross-Community Semantic Divergence

The following words were selected from the high-divergence list, based on comparisons between separately trained models for each community. We begin with words identified as having high semantic divergence across communities; a complementary diachronic (temporal) analysis is presented in the following section.

- Because some neighbors are shared between the two sets, we inspect the top 30 neighbors for each instance of the word.

<table>
<tr><th colspan="2">1. lefty / lefties / libs (cluster) shift = 0.465, 0.440, 0.355</th></tr>
<tr><td><u>Partial right-side neighbors:</u>

far-leftist, feminazi, leftard, leftoid, librul, libtarded, shitlib, progressive, ultra-left, commie, libtard, sjw</td><td><u>Partial Left-side neighbors:</u>

liberal, libertarian, socialist, marxist, neoliberals, conservatives, centrists, hippies, die-hard</td></tr>
<tr><td colspan="2">What stands out immediately is the tone difference.
On the right side, many neighbors are clearly constructed insults:
• libtard / libtarded → attaches a cognitive slur to “liberal”
• shitlib → explicit vulgar dismissal
• leftoid / leftard → dehumanizing suffixes
• feminazi → equates feminism with extremism
These are not neutral descriptors—they are labels designed to dismiss a group.</td></tr>
</table>

2. SJWs (Social Justice Warriors) shift = 0.376

Partial Right-side neighbors:	**Partial Left-side neighbors:**
blue-haired, feminazis, soyboys, woketards, wokies, wokesters, lunatics, morons, shitlibs, marxists, progressives, karens, commies	*conservatives, right-wingers, nationalists, homophobes, misogynists, radicals, trumpers, socialists*

The right-side neighbors are highly distinctive and consistent.
A few examples:

- *soyboy* → insult implying weakness or lack of masculinity
- *woketard / wokies* → merges "woke" with a slur
- *blue-haired* → stereotype of progressive activists

These are stereotypes.
On the left side, the neighbors do not form a parallel structure around "SJW". Instead, they refer to opposing groups (*conservatives, right-wingers, nationalists*), also, it seems that the term appears in meta-discussion / counter-labeling.

3. MAGA (Make America Great Again) shift = 0.313

Partial Right-side neighbors:	**Partial Left-side neighbors:**
beanie, camo, merch, shirt, t-shirt, tshirts, swag, wearing, trump2020, trump/pence.	*cult, cultist, cultists, inbred, loon, loons, incel, zombies, qanon, neo-nazis, trumpers*

This example is especially clear because the neighbors refer to very different types of things.
On the right:
hat, beanie, camo, t-shirt, merch, swag
These are physical items—objects people wear.
"MAGA" appears as a visible identity, something you display.
On the left:
cult, cultist, zombies, inbred, incel, qanon
These are strong negative characterizations**.**
The contrast:

- right → identity as something you wear and belong to.
- left → identity as something irrational or dangerous.

4. rioters / rioting / looting (cluster) shift = 0.331, 0.307, 0.304

Partial Right-side neighbors:	**Partial Left-side neighbors:**
looters, thugs, criminals, vandals, arsonists, lawlessness, mayhem, rioting/looting, agitators, anitfa, trespassers, vandals, terrorizing, blm/antifa	*1/6, capitol, j6, storming, siege, attackers, traitors, gassed, protester, brutality, non-violent, cops.*

This example is important because it shows different event anchoring.
On the right:
riots are associated with lawlessness and destruction, the term is also tied to Black Lives Matter protests (framed as criminal violence), general criminal framing.
On the left:
These refer to the January 6 Capitol event, a specific political incident (framed as political extremism).
Both are producing divergent meanings.

5. rightwing shift = 0.327	
<u>Partial Right-side neighbors:</u> left-leaning, leftist, leftwing, right-leaning, non-leftist, fringe, conservative, nationalist	**<u>Partial Left-side neighbors:</u>** *anti-democratic, anti-government, conspiratorial, fearmongering, brainless, incel, wingnut, anti-american, crazed, diarrhea, trumpie, trumpies*
On the right: The neighbors are mostly positional (*left-leaning, right-leaning, fringe*), describing where something sits on the spectrum. On the left: the neighbors shift to evaluation and accusation.	

6. progressive shift = 0.321	
<u>Partial Right-side neighbors:</u> *collectivist, marxist, radical, socialist/communist, statist, regressive*	**<u>Partial Left-side neighbors:</u>** *pragmatic, centrist, populist, sanders, warren, electable, anti-war, bernies, idealist, likeable*
On the right: The term is linked to ideological extremes. On the left: It appears within internal political discussion: (*sanders, warren* (politicians)), *pragmatic, electable* (strategy-oriented terms).	

7. traitor shift = 0.32643884	
<u>Partial Right-side neighbors:</u> *sellout, turncoat, deserter, coward, rino, warmonger, flip-flopper, lier, loser, pretender*	**<u>Partial Left-side neighbors:</u>** *grifter, psychopath, tyrant, thief, seditious, conman*
Both sides use strong language here, but with slightly different emphasis: Right: inter group " betrayal" - betrayal of group loyalty. RINO=(Republican In Name Only), Flip-flopper: Describes a politician who changes their position on an issue frequently, Sellout: Someone who abandons their principles. Left: corruption, abuse of power.	

8. transgender shift = 0.303	
<u>Partial Right-side neighbors:</u> *tranny, transgenderism, transpeople, non-binary, mtf, ftm*	**<u>Partial Left-side neighbors:</u>** *affirming, discrimination, transitioning, hormone, therapy, bathrooms*

The right focus on categorization and terminology (sometimes derogatory (highly offensive), e.g., *tranny*), Transgenderism - Supporters of trans rights rarely use it. The left focus on processes and rights: (transitioning, therapy, discrimination).

9. activist / activists 0.268 , 0.298	
Partial Right-side neighbors: *agitator, extremist, militant, ideologue, feminazi, sjw, pro-lgbt, pro-trans*	**Partial Left-side neighbors:** *organizer, founder, columnist, legislator, naacp*
On the right: activist is associated with: • *agitator, extremist, militant* → disruptive or radical action = destabilizing actor On the left: activist connects to: • *organizer, founder, institutions (NAACP)* → structured and legitimate roles = legitimate participant in civic life	

9.2 Temporal Semantic Divergence

The following examples were selected from temporal analyses, including high trend scores (with Z scores) and substantial early–late differences (with Δshift(w)). For each word, neighbors were examined across four models: left/right and early/late periods:

- ❖ We inspect the top 20 neighbors for each instance of the word (and from each era).

1. china Δshift(w) = 0.057
ERA A neighbors: *[belarus, iran, Nepal, phillipines, tibet, india, japan, korea, Taiwan], trade, asia, economy* **ERA B neighbors:** *[several countries], ccp, wuhan, virus, chicoms, russiachina, xi* In the earlier period, *china* appears in a standard geopolitical context: • *countries, trade, economy* → neutral or descriptive In the later period, it seems that the neighbors shift toward conflict and tension: • *ccp* (Chinese Communist Party) • *wuhan, virus* (COVID-19 origin discussions) • *chicoms* → slang, often used derogatorily to refer to Chinese communists
2. joe (Biden) Z_score = 2.024 **ERA A neighbors:** *joe, biden, democrats*

ERA B neighbors:
sleepyjoe, corrupt, illegitimate, senile, dementiaaddled, gropey, sleepy, dino

In the earlier period, references are mostly standard political mentions.
In the later period, we see:

- *sleepyjoe* → nickname used to portray weakness or cognitive decline
- *corrupt, illegitimate* → direct delegitimization

This reflects a shift toward personalized political language:

- less about policies
- more about character framing and attack labels

3. harris (Kamala Harris) Z_score = 1.980

ERA A neighbors:
shultz, tulsi, cory, cuomo, gavin, sanchez, Schweitzer

ERA B neighbors:
Variations and wordplays on the name which were not added (censored), as they were too vulgar.

We infer that early words were "neutral political references", whereas, the later neighbors include insults that target appearance or gender (gendered delegitimization)

4. fraud Δshift(w) = 0.029

ERA A neighbors:
fraud, corruption, scams, ponzi

ERA B neighbors:
election, ballots, voting, cheating, mailinvoting, voter, voterelection, voterfraud, tampered

Earlier:

- general concept → financial or legal wrongdoing

Later:

- tightly connected to elections distrust.

5. woke Δshift(w) = 0.057

ERA A neighbors:
woke, waking, awakened

ERA B neighbors:
wokeism, sjw, politically correct, woketards, wokies

Earlier:

- literal meaning → related to being awake

Later:

- completely different usage:
    - *wokeism* → ideology
    - *sjw* → social justice label
    - *woketards, wokies* → derogatory forms

Although the term has existed since 2000, it has undergone a drastic and full semantic transformation over time (due to certain events).

6. nursing Δshift(w) = 0.070

ERA A neighbors:
icu, hospital, rotations, therapists, assistants

ERA B neighbors:
covid, nursinghome, elderly, cuomo, cumo, patients, homes, seniors, vunerable, basements, overflow, quarantining

Earlier:

- clearly a medical profession context

Later:

- strongly tied to COVID:
  - *nursinghome, elderly* → high-risk populations
  - *cuomo* → political controversy related to policy decisions of the NY governor.

It appears that there has been a shift in framing from profession-related concepts toward political blame and tension.

7. terrorists Δshift(w) = 0.039

ERA A neighbors:
isis, taliban, jihadists, afghanis, militants, Hezbollah, isis,

ERA B neighbors:
rioters, insurrectionists, extremists, seditionists, antifia, subversives, looters, facists

Earlier:

- clearly external actors (foreign threat):
  - *ISIS, Taliban*

Later:

- includes domestic actors (internal political labeling):
  - *rioters, insurrectionists*

Across the examined examples, several recurring patterns emerge. One prominent pattern is **asymmetric labeling**, where a term is used descriptively by one community but as a **derogatory** label by the other. More broadly, the observed differences often reflect variation not only in topic, but also in tone, intent, and rhetorical function. In some cases, divergence appears as a distinction between **identity-based framing** and moral or evaluative framing. In others, it is linked to specific **political events**, with different communities associating the same term with different real-world contexts.

It is important to acknowledge that, in some instances, what is measured as semantic divergence may partly result from differences in contextual usage rather than deep ideological shifts. That is, certain distinctions may be lexical or stylistic rather than fundamentally conceptual. Nevertheless, the overall volume of terms exhibiting strong divergence suggests a broader and **systematic phenomenon**. In particular, the findings indicate that, over time, different communities increasingly use language in distinct ways—**an observation that aligns with the research questions concerning political polarization**.

9.3 Conclusions and Implications

Finally, the qualitative findings provide important support for the quantitative results. The presence of numerous clear and interpretable examples indicates that the measured semantic shifts are not merely artifacts of the modeling process. Rather, they correspond to recognizable and meaningful differences in language use, both across ideological groups and over time. **Taken together, these results strengthen the overall conclusion** that semantic divergence is a substantive and measurable aspect of political polarization.

More broadly, the findings of this work address the central research questions by demonstrating that the semantic meaning of political terms **is not static, but evolves both within and across ideological communities.** The observed differences between left- and right-leaning corpora reveal consistent patterns of divergence, separation, and drift in meaning. Furthermore, **the temporal analysis shows that these gaps are not fixed, but tend to expand over time.**

While this phenomenon alone does not fully capture the complexity of political polarization, the results suggest that semantic divergence constitutes one observable and measurable aspect of it. In this sense, the study provides empirical support for the claim that **polarization is reflected not only in opinions or attitudes, but also in the structure and usage of <u>language</u> itself.**

10. Future Work

Several directions can extend and strengthen the present study.

First, alternative embedding models such as FastText could be explored to evaluate whether subword information improves robustness, particularly for rare or morphologically complex terms. Comparing results across models would provide additional validation of the findings.

Second, a global quantitative metric—such as an Aggregate Polarization Index (API)—could be developed to summarize overall semantic divergence across a predefined set of politically relevant terms. This would enable easier comparison across datasets and time periods.

Third, expanding the dataset to include more recent years would be particularly valuable. Extending the analysis beyond 2023 would enable examination of more recent global developments characterized by heightened political, economic, and social tensions. Applying the same analytical framework to such data could help assess whether semantic polarization continues to intensify under these conditions.

Additional analyses could also be conducted on the structure of the embedding space. Clustering techniques, combined with dimensionality reduction methods such as t-SNE or UMAP, could be used to visualize how ideological language separates over time. Similarly, tracking changes in nearest-neighbor structures could provide a more detailed view of how specific terms shift in meaning.

Finally, more advanced contextual models such as Sentence-BERT could be used to move beyond word-level analysis and capture polarization at the sentence or discourse level. This would allow examination of whether entire narratives or topics become more polarized, complementing the lexical analysis presented in this work.

10. References

1. **Goldin, G., Rabinovich, E., & Wintner, S. (2025). Unveiling Affective Polarization Trends in Parliamentary Proceedings.**

2. Jensen, Jacob, Ethan Kaplan, Suresh Naidu, and Laurence Wilse-Samson. 2012. Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech. Brookings Papers on Economic Activity, 43(2 (Fall)):1–81.

3. Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Jesse Shapiro, Matthew Gentzkow, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. 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 2970–3005, Association for Computational Linguistics, Minneapolis, Minnesota.

4. Milbauer, Jeremiah, Adarsh Mathew, and James Evans. 2021. Aligning multidimensional worldviews and discovering ideological differences. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 4832–4845, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic.

5. Sinno, Barea, Bernardo Oviedo, Katherine Atwell, Malihe Alikhani, and Junyi Jessy Li. 2022. Political ideology and polarization: A multi-dimensional approach. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 231–243, Association for Computational Linguistics, Seattle, United States.

6. Frimer, Jeremy A., Harinder Aujla, Matthew Feinberg, Linda J. Skitka, Karl Aquino, Johannes C. Eichstaedt, and Robb Willer. 2022. Incivility is rising among American politicians on Twitter. Social Psychological and Personality Science.

7. Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. Affect, not ideology: A social identity perspective on polarization. Public Opinion Quarterly, 76(3):405–431.

8. Mehrabian, Albert and James A. Russell. 1974. An Approach to Environmental Psychology. MIT Press, Cambridge, MA.

9. **Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.**

10. **Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems* (pp. 3111–3119).**

11. Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*(1), 1–10.

12. **Hill, F., Reichart, R., & Korhonen, A. (2015). SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation. Computational Linguistics, 41(4), 665–695.**

13. **Monroe, Colaresi, and Quinn (2008) in *Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict***

14. Shem-Tov Eliran, Ella Rabinovich, (2025). *On the interplay between musical preferences and personality through the lens of language*. In Proceedings of the CEUR Workshop (Vol. 4136). https://ceur-ws.org/Vol-4136/iaai8.pdf

15. Preoțiuc-Pietro, D., Liu, Y., Hopkins, D., & Ungar, L. (2017). *Beyond binary labels: Political ideology prediction of Twitter users*. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), 729–740.

16. Rabinovich, Ella, Yulia Tsvetkov, and Shuly Wintner.
Native Language Cognate Effects on Second Language Lexical Choice.
In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1790–1801, 2018.

17. Pratapa, Adithya, Monojit Choudhury, and Kalika Bali.
CodeSwitch-Reddit: Exploration of Written Multilingual Discourse in Online Discussion Forums.
In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*, 2018.

18. Schneider, Nathan, Yulia Tsvetkov, and others.
Pick a Fight or Bite your Tongue: Investigation of Gender Differences in Idiomatic Language Usage.
In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

19. Mann, H. B. (1945). *Nonparametric tests against trend*. Econometrica, 13(3), 245–259.

20. Kendall, M. G. (1975). *Rank Correlation Methods* (4th ed.). Charles Griffin.

12. Appendix

A.

Basic descriptive statistics:

- The source below is the "Right Wing" data:

Name:	r.conservative (posts + comments)	
Total records:	11,333,576	
Date:	Earliest:	2008-03-06 02:39:23
	Latest:	2023-04-09 18:35:11
Total sentences:	30,177,471	

Average = (sentences / records):	2.66	
Sentence length (by words):	Mean:	13.38
	Median:	11.00
	Max:	4894

Name:	r.democrats (posts + comments)	
Total records:	843,287	
Date:	Earliest:	2009-02-16 13:31:39
	Latest:	2023-04-12 07:14:15
Total sentences:	2,392,289	
Average = (sentences / records):	2.84	
Sentence length (by words):	Mean:	13.30
	Median:	11.00
	Max:	1286

Name:	r.liberal (posts + comments)	
Total records:	298,355	
Date:	Earliest:	2009-05-03 03:04:24
	Latest:	2023-04-12 11:47:56
Total sentences:	989,280	
Average = (sentences / records):	3.32	
Sentence length (by words):	Mean:	14.17
	Median:	12.00
	Max:	1455

- The source below is the "Left Wing" combined data:

Name:	Left (r.liberal + r.democrats)	
Total records:	1,141,642	
Date:	Earliest:	2009-02-16 13:31:39
	Latest:	2023-04-12 11:47:56
Total sentences:	3,381,569	
Average = (sentences / records):	2.96	
Sentence length (by words):	Mean:	13.56
	Median:	11.00
	Max:	1455

#	Time:	Info	
1	Before 2010	📊 Combined CSV Analysis (left_split_1.csv) Total records: 26 **Total sentences: 107** Mean sentences per record: 4.12 Sentence length (by words): Mean: 14.18 Median: 12.00 Max: 60	📊 Combined CSV Analysis (r.conservative_split_1.csv) Total records: 962 **Total sentences: 3099** Mean sentences per record: 3.22 Sentence length (by words): Mean: 13.85 Median: 11.00 Max: 237
2	2010-2011	📊 Combined CSV Analysis (left_split_2.csv) Total records: 188 **Total sentences: 1015** Mean sentences per record: 5.40 Sentence length (by words): Mean: 14.97 Median: 12.00 Max: 113	📊 Combined CSV Analysis (r.conservative_split_2.csv) Total records: 4954 **Total sentences: 16568** Mean sentences per record: 3.34 Sentence length (by words): Mean: 14.88 Median: 12.00 Max: 115
3	2011-2012	📊 Combined CSV Analysis (left_split_3.csv) Total records: 6031 **Total sentences: 21743** Mean sentences per record: 3.61 Sentence length (by words): Mean: 15.26 Median: 13.00 Max: 467	📊 Combined CSV Analysis (r.conservative_split_3.csv) Total records: 40881 **Total sentences: 154338** Mean sentences per record: 3.78 Sentence length (by words): Mean: 14.74 Median: 12.00 Max: 282
4	2012-2013	📊 Combined CSV Analysis (left_split_4.csv) Total records: 23442 **Total sentences: 86403** Mean sentences per record: 3.69 Sentence length (by words): Mean: 15.13 Median: 13.00 Max: 504	📊 Combined CSV Analysis (r.conservative_split_4.csv) Total records: 228730 **Total sentences: 802558** Mean sentences per record: 3.51 Sentence length (by words): Mean: 14.85 Median: 13.00 Max: 544
5	2013-2014	📊 Combined CSV Analysis (left_split_5.csv) Total records: 36225 **Total sentences: 139401** Mean sentences per record: 3.85 Sentence length (by words): Mean: 14.90 Median: 12.00 Max: 800	📊 Combined CSV Analysis (r.conservative_split_5.csv) Total records: 160297 **Total sentences: 516003** Mean sentences per record: 3.22 Sentence length (by words): Mean: 14.67 Median: 12.00 Max: 568
6	2014-2015	📊 Combined CSV Analysis (left_split_6.csv) Total records: 36543 **Total sentences: 140678** Mean sentences per record: 3.85 Sentence length (by words): Mean: 14.75 Median: 12.00 Max: 834	📊 Combined CSV Analysis (r.conservative_split_6.csv) Total records: 153960 **Total sentences: 505878** Mean sentences per record: 3.29 Sentence length (by words): Mean: 14.68 Median: 12.00 Max: 701
7	2015-2016	📊 Combined CSV Analysis (left_split_7.csv) Total records: 43084 **Total sentences: 153600** Mean sentences per record: 3.57 Sentence length (by words): Mean: 14.83 Median: 12.00 Max: 1188	📊 Combined CSV Analysis (r.conservative_split_7.csv) Total records: 201343 **Total sentences: 619972** Mean sentences per record: 3.08 Sentence length (by words): Mean: 14.35 Median: 12.00 Max: 628

8	2016-2017	📊 Combined CSV Analysis (left_split_8.csv) Total records: 90029 **Total sentences: 319696** Mean sentences per record: 3.55 Sentence length (by words): Mean: 14.05 Median: 12.00 Max: 1455	📊 Combined CSV Analysis (r.conservative_split_8.csv) Total records: 479518 **Total sentences: 1461985** Mean sentences per record: 3.05 Sentence length (by words): Mean: 13.97 Median: 12.00 Max: 763
9	2017-2018	📊 Combined CSV Analysis (left_split_9.csv) Total records: 68054 **Total sentences: 229851** Mean sentences per record: 3.38 Sentence length (by words): Mean: 14.47 Median: 12.00 Max: 1259	📊 Combined CSV Analysis (r.conservative_split_9.csv) Total records: 547454 **Total sentences: 1681458** Mean sentences per record: 3.07 Sentence length (by words): Mean: 14.47 Median: 12.00 Max: 4894
10	2018-2019	📊 Combined CSV Analysis (left_split_10.csv) Total records: 114899 **Total sentences: 353426** Mean sentences per record: 3.08 Sentence length (by words): Mean: 13.80 Median: 11.00 Max: 1015	📊 Combined CSV Analysis (r.conservative_split_10.csv) Total records: 521878 **Total sentences: 1446534** Mean sentences per record: 2.77 Sentence length (by words): Mean: 13.92 Median: 12.00 Max: 1228
11	2019-2020	📊 Combined CSV Analysis (left_split_11.csv) Total records: 109156 **Total sentences: 306862** Mean sentences per record: 2.81 Sentence length (by words): Mean: 13.50 Median: 11.00 Max: 1011	📊 Combined CSV Analysis (r.conservative_split_11.csv) Total records: 776934 **Total sentences: 2081886** Mean sentences per record: 2.68 Sentence length (by words): Mean: 13.69 Median: 11.00 Max: 2736
12	2020-2021	📊 Combined CSV Analysis (left_split_12.csv) Total records: 185803 **Total sentences: 504439** Mean sentences per record: 2.71 Sentence length (by words): Mean: 13.05 Median: 11.00 Max: 1286	📊 Combined CSV Analysis (r.conservative_split_12.csv) Total records: 3103956 **Total sentences: 8382320** Mean sentences per record: 2.70 Sentence length (by words): Mean: 13.30 Median: 11.00 Max: 846
13	2021-2022	📊 Combined CSV Analysis (left_split_13.csv) Total records: 180568 **Total sentences: 493491** Mean sentences per record: 2.73 Sentence length (by words): Mean: 12.80 Median: 10.00 Max: 660	📊 Combined CSV Analysis (r.conservative_split_13.csv) Total records: 2723634 **Total sentences: 6946863** Mean sentences per record: 2.55 Sentence length (by words): Mean: 13.04 Median: 11.00 Max: 1015
14	After 2022	📊 Combined CSV Analysis (left_split_14.csv) Total records: 51104 **Total sentences: 122840** Mean sentences per record: 2.40 Sentence length (by words): Mean: 12.42 Median: 10.00 Max: 319	📊 Combined CSV Analysis (r.conservative_split_14.csv) Total records: 372957 **Total sentences: 876459** Mean sentences per record: 2.35 Sentence length (by words): Mean: 12.84 Median: 11.00 Max: 439

B.

Hyperparameters specifications:

min_count=100:

Ignore any word that appears fewer than *min_count*. This dramatically shrinks the vocabulary. Rare words are noisy and not worth allocating a standard 300-dimensional vector.

window=2:

Word2Vec uses a sliding window around a target word. Window size 2 means: look at 2 words to the left and 2 words to the right. Smaller windows focus more on syntactic similarity ("comes before", "modifier of"). Larger windows (4–8) capture more semantic similarity. Window 2 is considered a tight, grammar-ish window.

vector_size=300:

Dimensionality of word embeddings. Higher = more expressive.

sample=6e-5:

Subsampling of frequent words. Words like "the", "you", "it" appears so often they drown out learning. Subsampling randomly discards some of their occurrences.

negative=20:

Negative sampling uses 20 negative examples for each positive example. Higher = better quality but slower.

workers=cores-1:

Parallelism. On big corpora, this speeds up training massively.

Table B.1 — Spearman Correlation (SimLex-999):

Vector Size	Window	Min Count	Subsampling	Negative Sampling	Spearman Correlation
300	**2**	**100**	**6e-05**	**20**	**0.3963**
300	3	100	6e-05	20	0.3848
200	2	100	6e-05	20	0.3820
300	4	100	6e-05	20	0.3739
200	3	100	6e-05	20	0.3668
300	5	100	6e-05	20	0.3533
200	4	100	6e-05	20	0.3501
200	5	100	6e-05	20	0.3384

Vector Size	Window	Min Count	Subsampling	Negative Sampling	Spearman Correlation
100	2	100	6e-05	20	0.3382
100	3	100	6e-05	20	0.3168
100	4	100	6e-05	20	0.3061
100	5	100	6e-05	20	0.2923

We were fixed of min-count, subsampling, negative sampling.

Table B.2 — Pearson Correlation (SimLex-999):

Vector Size	Window	Min Count	Subsampling	Negative Sampling	Pearson Correlation
300	2	100	6e-05	20	**0.4206**
300	3	100	6e-05	20	0.4049

C.

Partial list of the final politically relevant terms (180 from 851):

trump	law	leftists	nation	million
biden	votes	people	countries	schools
vote	rights	economy	border	florida
democrats	war	racism	president.	government.
government	democratic	public	health	fbi
conservative	trump.	violence	wing	businesses
president	voters	illegal	children	communist
election	americans	laws	socialist	conservative.
political	politics	cnn	death	ballot
left	evidence	healthcare	history	pelosi
republicans	leftist	fraud	romney	welfare
state	court	majority	socialism	party.
conservatives	abortion	country.	population	attack
country	taxes	administration	freedom	constitutional
republican	senate	virus	climate	pandemic
liberal	vaccine	politicians	vote.	racist.
obama	china	election.	facts	governor
states	hillary	citizens	bush	progressive
media	military	fox	antifa	riots
party	russia	civil	narrative	criminal
us	voter	supporters	legal	defend
covid	win	elected	texas	income
democrat	policies	supreme	ukraine	vaccinated
tax	guns	deaths	murder	sanders

racist	literally	justice	immigrants	cities
voting	cruz	immigration	candidates	biden.
black	candidate	united	ban	protest
liberals	policy	ballots	foreign	cops
police	race	amendment	desantis	crimes
gun	crime	economic	russian	disagree
american	constitution	california	speech	democrats.
federal	congress	bill	polls	dem
dems	campaign	won	violent	religious
america	trump's	right	donald	democracy
news	clinton	agree	education	propaganda
gop	bernie	elections	power	banned

D.

Partial list of words by semantic shift (34 from 851):

word	cosine_similarity_R_vs_D	semantic_shift
corona	0.5121911	0.48780888
discord	0.5125213	0.48747867
lefty	0.53433526	0.46566474
appears	0.5546146	0.4453854
neocons	0.55752826	0.44247174
lefties	0.55958426	0.44041574
maxine	0.5674614	0.43253863
senile	0.5681868	0.43181318
radicals	0.5755013	0.42449868
vaccinations	0.5786226	0.42137742
marxist	0.58056897	0.41943103
ccp	0.58292973	0.41707027
regarding	0.58717763	0.41282237
mayors	0.587392	0.41260803
mandates	0.5905554	0.40944457
dominion	0.59653544	0.40346456
historic	0.596642	0.40335798
schiff	0.60211945	0.39788055
lefts	0.61662066	0.38337934
mobs	0.6189647	0.38103533
leftist	0.61972153	0.38027847
sjws	0.62304676	0.37695324
trump	0.62439317	0.37560683
abrams	0.6291215	0.37087852
powell	0.6306716	0.36932838
brigade	0.63126326	0.36873674
lockdown	0.6326864	0.36731362
farleft	0.6330288	0.3669712

taiwan	0.6354527	0.3645473
admin	0.6359062	0.36409378
woke	0.6366382	0.36336178
pedophiles	0.6369593	0.3630407
prochoice	0.63750935	0.36249065
giuliani	0.6382468	0.36175323