

Article

Not peer-reviewed version

Mapping the Semantic Networks of Political Communication: Diachronic Transitions from Structurally Coherent to Semantically Fragmented Discourse in the Digital Era

[Sophia Melanson Ricciardone](#) *

Posted Date: 30 April 2026

doi: 10.20944/preprints202604.2083.v1

Keywords: semantic networks; political discourse; semantic coupling; discourse fragmentation; corpus linguistics; computational discourse analysis; political polarization; media ecology; institutional trust; diachronic analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Mapping the Semantic Networks of Political Communication: Diachronic Transitions from Structurally Coherent to Semantically Fragmented Discourse in the Digital Era

Sophia Melanson Ricciardone

McMaster University; melans1@mcmaster.ca

Abstract

This study examines how the semantic features of political discourse have changed since the pre-digital period and how contemporary political news media environments have become structurally organized according to fragmented meaning ecologies. Rather than examining lexical content alone, this study uses a comparative diachronic semantic network analysis using corpus-based computational discourse methods. **Background:** While polarization is often studied in terms of ideological distance or word frequency shifts, less is known about how the relationships between semantic domains themselves may reorganize over time, potentially affecting social cohesion and institutional trust. **Methods:** A comparative diachronic design was applied to political news transcripts from the 1980s and the 2020's (both sourced via YouTube). Semantic annotation (WMatrix7), n-gram analysis, and Pearson correlation-based semantic network modeling were used to compare semantic coupling across governance, emotional, psychological, and social domains, alongside distributional statistics and functional discourse coding. **Results:** independent t-tests found no significant differences were found in overall semantic frequency distributions between corpora, indicating distributional stability. However, network analyses revealed a strong contrast in structure: the 1980 corpus exhibited uniformly strong positive correlations across semantic domains, reflecting a highly integrated system, whereas the 2026 corpus showed weaker, more variable, and in some cases negative correlations, strongly suggesting semantic decoupling and fragmentation. These findings were supported by n-gram and functional analyses showing increased conversationalization, negation, and disfluency in the 2026 discourse. **Conclusions:** Political discourse appears to have undergone a structural reorganization from a coherent, highly coupled semantic system toward a more modular and fragmented configuration, suggesting that contemporary polarization is better understood as a shift in semantic organization than as changes in lexical frequency alone.

Keywords: semantic networks; political discourse; semantic coupling; discourse fragmentation; corpus linguistics; computational discourse analysis; political polarization; media ecology; institutional trust; diachronic analysis

1. Introduction

Dating as far back as The Republic, politics has long been an arena for the cultivation of contentious discourse, both among politicians and within the public sphere. At the same time, contemporary political debate has not only become more contentious but increasingly divisive and polarized at both the political and societal levels. Indeed, in this third decade of the twenty first century, the boundaries that enshrine political in-group and out-group distinctions have assumed a distinctly partisan character. Though a range of social, economic, ecological, and global pressures most certainly contribute to shifts in political culture, one significant technological advancement in

media communication that has gained notable prominence for its pervasive influence across nearly every aspect of collective life is the rise of the Internet and the subsequent emergence of artificial intelligence (AI) technologies.

In the United States, citizens are now nearly completely divided along ideological lines, with 92% of Republicans positioned to the right of the median and 94% of Democrats to the left, leaving very few occupying the center (Pew Research Center, 2014 [1]— see Figure 1). This reflects more than mere disagreement; it implies substantial erosion of ideological overlap, with the middle ground effectively collapsing on itself. This degree of polarization not only compromises the integrity of political cooperation across democratic societies but cultivates fragmentation of the public sphere as the structure of shared narratives breaks down, creating cultural dislocation in the process. Such dislocation likewise weakens normative authority and contributes to anomic (Durkheim, 2002 / 1897 [2]; Merton, 1938 [3]) conditions—the condition where socially shared norms lose their clarity, authority, or effectiveness, leaving individuals without stable guidance for conventions on how to think, act, and belong in a socially common space with others.

Advancements in digitally mediated communication and human-AI interaction has transformed how social and political discourse unfolds within the public sphere. Increasingly, communication takes place within algorithmically mediated environments, where AI systems, including recommendation algorithms, chatbots, and generative AI tools regulate the flow of content and determine which political narratives become amplified, suppressed, or validated and which are occluded. Unlike traditional mass media, these systems are dynamic, interactive, personalized, and highly networked, generating emergent patterns of meaning-making that depend on perpetual recursive engagement between algorithmic processes and user behavior within online spaces. This interaction creates conditions where ideas, frames, and norms can propagate instantaneously in semantically ‘contagious’ ways across digital networks (Gündüz, 2017 [4]; Jordan, 2020 [5]; Rehman, 2025[6]), reconceptualizing the dynamics of social groups and institutions.

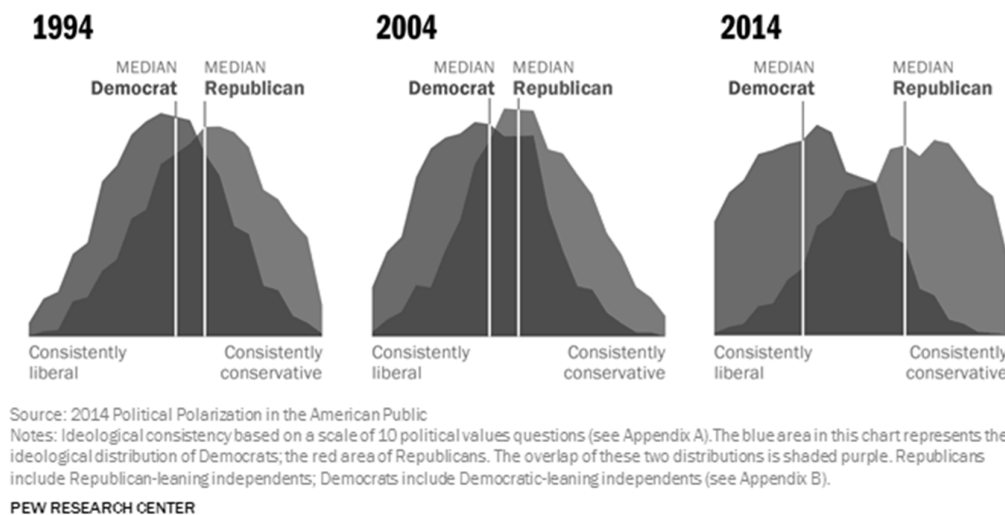


Figure 1. Distribution of Democrats and Republicans on a 10-item scale of political values.

Despite extensive research on political polarization, most content analysis studies have focused on shifts in word frequencies, sentiment, or ideological positioning over time, privileging surface-level variation while overlooking the relational structure through which meaning is produced and maintained in discourse (Abercrombie & Batista-Navarro, 2020 [7]; Németh, 2022 [8]; Haselmayer & Jenny, 2017 [9]). While informative, these approaches tend to treat political language as a collection of discrete elements rather than as a structured system of interactions, overlooking how meaning emerges from patterns of contextual and relational use (Farzindar & Inkpen, 2020 [10]; Camacho-

Collados et al., 2022 [11]). In consequence, they fail to capture how meanings cohere or fail to cohere semantically across discourse.

Accordingly, this study demonstrates how digital transformations to contemporary political communication are not merely lexical or affective, but fundamentally structural, shifting away from a relatively integrated semantic mediating systems (e.g., television, radio, newsprint) to one that weakly couples and partially decouples the semantic salience of political information in the public sphere. From a cognitive science perspective, the coherent integration of meaning depends on the consolidation of information across representational domains, such that breakdowns in this integration produce instability and fragmentation in the interpretation of meaning among constituents (Clark, 2013 [12]; Friston, 2010 [13]; Pickering & Garrod, 2004 [14]).

In digital media environments, this process is further disrupted by AI regulation of algorithmic curation and automated agents, which restructure the flow and synthesis of information while, at the same time, amplifying fragmentation across communicative networks (Ferrara et al., 2016 [15]; Stella et al., 2018 [16]; Mansoury et al., 2020 [17]; Woolley, 2020 [18]; Prior, 2007 [19]; Bennett & Iyenga, 2008 [20]; Klinger & Svensson, 2015 [21]; Van Aelst et al., 2017 [22]; Van Dijck, Poell & de Waal, 2018 [23]). In these communicative environments, relationships between semantic domains become attenuated, unstable, and fragmented, reducing the overall coherence and structure of meaning. Within this framework, polarization is not treated solely as ideological distance, but as a weakening of cross-domain semantic integration of networked information, whereby meanings that were previously co-articulated across institutional, social, and psychological domains become increasingly disaggregated, perpetuating a process further reinforced by motivated reasoning and heuristic-based judgment under conditions of uncertainty (Kahan, 2013 [24]; Tversky & Kahneman, 1974 [25]).

Conceptualizing discourse in terms of semantic coupling provides a means through which to make this transformation analytically tractable. In an integrated communication system, for instance, semantic categories are densely interconnected, with patterns of co-occurrence reflecting stable relationships between domains such as institutions, values, and social actors, which is consistent with the view that statistical learning systems capture underlying relational structure in language use (Bottou et al., 2016 [26]; Schmidhuber, 2014 [27]). By contrast, a decoupled system is characterized by weaker, more fragmented semantic associations, in which meanings circulate in more isolated or loosely connected semantic clusters and categories of meaning.

Accordingly, this study operationalizes polarization as a shift in the strength and structure of these relationships, allowing discursive fragmentation to be measured rather than inferred. The rise of digitally mediated communication, and more particularly, AI-mediated systems, provides an important context for this transformation. Political polarization predates these technologies, with identity-based sorting already accelerating in the 1990s–2000s (Mason, 2018 [28]). However, today, algorithmically mediated environments that adapt deep learning techniques now reorganize how content is semantically assembled and disseminated within discourse by shaping visibility, amplification, and interaction among human users. Unlike traditional unilateral media consumption configurations, such as television, radio, and print journalism, digital systems are interactive, personalized, and recursive, generating feedback loops between user behavior and algorithmic curation. As discourse increasingly unfolds within these segmented network environments (often described as echo chambers and filter bubbles), exposure to meaning becomes unevenly distributed and this fragmentation becomes continuously reinforced (Hedayatifar et al., 2019 [29]).

From this point of reference, the fragmentation of discourse can be understood as an emergent property of these communicative conditions. When semantic relationships are repeatedly reinforced within bounded networks while remaining weakly interconnected across them, the overall structure of discourse may shift toward partial decoupling. This does not simply produce stronger opinions; it alters the organization of meaning construction itself. Processes often described in terms of identity fragmentation, declining normative consensus, and institutional distrust can thus be interpreted as downstream implications of this structural reconfiguration (Kossowska et al., 2023 [30]). In classical sociological terms, such conditions resemble forms of anomie, in which shared norms and

expectations become destabilized among citizens (Durkheim, 2002 / 1897 [2]; Merton, 1938 [3]; Rehman, 2025 [6]), though here this instability is observable at the level of meaning and discourse rather than solely at the level of behavior.

For these reasons, this study examines whether the semantic dynamics of political news discourse in the 1980s compared to the 2020s reflect this structural transformation. It does so by comparing patterns of semantic association in contemporary news discourse with those of pre-digital news media from the 1980s. Drawing on computational corpus methods, the analysis models relationships between semantic categories to assess the degree of coupling and decoupling that transpires within each period. In doing so, the study advances a new approach to studying political polarization that treats it as a problem of semantic organization rather than solely as ideological distance or affect. What follows is organized around the following research question, which asks:

RQ: Do the semantic dynamics of political news discourse reflect greater fragmentation and identity dislocation than those observed in pre-digital, unilaterally transmitted media environments of the 1980s?

The following hypotheses will be explored in responding to this question:

H1: Semantic Boundary Volatility (Micro-level): Semantic networks of political news discourse from the 2020s (collected via transcripts of YouTube videos) will exhibit weaker coupling, greater levels of uncertainty, and increased fragmentation, reflected in shifting referents and reduced category coherence, compared to traditional political discourse in news media from the 1980s.

H2: Semantic Contagion and Identity Diffusion (Meso-level): transcripts of 2020s news media will show increased semantic contagion, which will correspond to weaker coupling among identity-related terms, reflecting more diffuse and less coherent representations of group identity at the collective level.

H3: Normative Ambiguity Encoding (System-level): Discourse environments characterized by weakened semantic coupling will exhibit normative networks that are denser but less hierarchically structured, indicating reduced coherence in the organization of norms.

H4: Institutional Semantic Erosion (Macro-level): Semantic networks derived from the 2020s corpus will show weaker coupling between institutional actors and legitimacy-related constructs, reflecting erosion in the semantic foundations of institutional trust.

2. Materials and Methods

This study employed a comparative diachronic design to examine structural changes in political discourse across two media environments: pre-digital, traditionally mediated political news discourse from the 1980s and contemporary, political news discourse during the digital era in the 2020s. The analysis focuses on whether shifts in communicative environments correspond with changes in the structure of semantic relationships, operationalized as variation in semantic coupling across domains.

2.1. Data and Corpus Construction

Two corpora were constructed for comparison: one consisting of political discourse drawn from archived news media broadcasts from the 1980s and another from contemporary news media broadcasts from the 2020s, both accessed via YouTube (see Appendix A for source details). The use of YouTube as a common retrieval platform serves a methodological purpose. By sourcing both historical and contemporary material from the same platform, the study controls for differences in

transcription format, accessibility, and mediation layer, thereby enhancing comparability across time periods and future accessibility of the data.

This analysis does not examine causal, correlational, or predictive relationships between digital media and political discourse, but rather, news media content is used as a longitudinal proxy for examining broader shifts in publicly circulating political news media language across historical periods, particularly in relation to changes in the wider media ecology. Accordingly, any observed differences between corpora are interpreted as indicating shifts in publicly mediated political discourse over time rather than as effects attributable to digital media, *per se*, in a direct way (though this could be achieved in future research). News media content was selected solely based on whether it contained significant reference to political issues relative to its corresponding period and was used as the basis for empirical comparison since it provided a consistent, archived, and publicly accessible record of political discourse across both periods studied. Although news media is not identical to everyday or digital-native communication, it serves as a standardized genre of political language production that is comparable across time, making it particularly suitable for diachronic linguistic analysis. Thus, this study does not test relationships between digital media and discourse change; instead, it uses temporal comparison of news discourse as a way of indexing broader semantic shifts in publicly available political language.

Importantly, the analysis does not treat YouTube as the object of study, but as a delivery infrastructure through which comparable broadcast content can be accessed, transcribed, and compiled for analysis. The 1980s corpus reflects discourse produced during a period dominated by traditionally mediated, unilaterally consumed communication, while the 2020s corpus reflects discourse produced during a period dominated by digitally mediated ecosystems characterized by algorithmic curation, interactivity, and human-AI interactions. Holding the mode of access constant while varying the underlying communicative environment allows the analysis to isolate diachronic changes in semantic structure, rather than artifacts introduced by differences in data collection or transcription. Each corpus was processed to ensure comparability in size and structure. Semantic tagging via WMatrix7 (Rayson, 2008 [31]) produced 127 comparable observations per corpus, representing normalized frequencies of semantic categories. The Youtube videos selected included archived CBC News clips from the 1980s and the 2020s, with the former transcripts (1980s) containing a total of 33,406 words and 150,736 characters and the latter (2020s) corpus containing 32,701 words and 149,742 characters.

2.2. Semantic Categorization

The textual data of the transcripts were semantically annotated using WMatrix7 (Rayson, 2008 [31]) a computational Natural Language Processing platform that groups lexical items into higher-order domains, including semantic tagging (semtags). Making operational use of the WMatrix7's built-in semantic categories and fields (GOVNT [governance, institutions, public authority], EMO [emotional states and expressions], PSYCH [psychological processes and internal states], and SOC [social relations, actions, and collective processes]), semtag frequencies were collected for each corpus. These categories served as the basis for modeling semantic networks, where relationships between domains were analyzed to reflect patterns of co-occurrence and discursive alignment.

2.3. Analytical Strategy

The analysis first tested whether the semantic structure of political discourse differs between a pre-digital (1980s) and digital (2020s) era. Examining semantic decoupling, the analysis proceeded in three stages involving a distribution analysis, a mean comparison, and a semantic network analysis. For the distribution analysis, descriptive statistics were computed for semtag frequencies across each corpus, including mean, variance, quartiles, and confidence intervals. Tests for normality were conducted using Kolmogorov-Smirnov, Shapiro-Wilk, and Anderson-Darling. These tests were used to determine whether parametric comparisons (e.g., independent sample t-tests) were appropriate for the analysis. Mean comparison was conducted using an independent samples t-test to assess

differences in overall semantic frequency distributions between the 1980 and 2026 corpora. Homogeneity of variance was tested using Levene's test and effect size was calculated using Cohen's *d*.

Finally, a semantic network analysis (the paper's core analysis) modeled discourse as a network of relationships between semantic categories, operationalized using *semtag* frequencies (e.g., GOVNT, EMO, PSYCH, and SOC) through Pearson correlation matrices. Correlations were computed separately for the 1980s and 2020s corpora across the political, emotional, psychological, and social semantic fields built into WMatrix7s infrastructure. Statistical significance was assessed for each pairwise relationship. Normality tests were conducted to validate correlation assumptions and matrices were interpreted as indicators of semantic coupling where higher scores represented consistent positive correlations and strong coupling (e.g., representing integrated system) and weaker scores represented variable or negative correlations as well as decoupling (e.g., fragmented system). Changes in correlation strength, direction (e.g., positive vs. negative), and overall coherence were used to assess structural and diachronic transformation in discourse between the 1980s and 2020s.

2.4. *N-gram Analysis for Discursive Patterning*

Complementing network-level analysis, a 4-gram (*n*-gram) analysis was conducted to examine phrase-level structure. For each corpus, frequencies (Freq) were collected for occurrence rate of *n*-grams along with Mutual Information (MI) scores to capture the strength of collocational binding between words co-occurring in ways not due to chance. Log-Likelihood (LL2) scores were also collected to account for the distinctiveness and strength of collocation in 4-grams (4-word coupling) relative to each corpus. This analysis allowed for the capture of formulaicity and linguistic stability (e.g., high MI score), discursive routinization in contrast to variability, and institutional and collective framing patterns. This step is intended to evaluate whether observed differences are attributable to overall distributional change as opposed to structural reorganization.

2.5. *Operationalization of Key Concepts*

Core theoretical constructs were operationalized according to the concepts of semantic coupling (e.g., to assess the strength and consistency of correlations between semantic categories); semantic decoupling (e.g., reduction in correlation strength, increased variance, or emergence of negative relationships); discursive fragmentation (e.g., system-level pattern characterized by weak, inconsistent, or oppositional semantic relationships); institutional anchoring (e.g., strength of associations between governance-related terms and broader semantic domains); and identity coherence (e.g., degree of alignment among social, emotional, and psychological semantic categories). Significance level was set at $p < .05$ and effect sizes were interpreted using conventional benchmarks for Cohen's *d*. Correlation strength was interpreted using standard thresholds (e.g., weak, moderate, strong) alongside *p*-values to measure the significance of correlation outcomes. By combining distributional statistics, inferential testing, and semantic network modeling, this study distinguishes between surface-level similarity (e.g., stable frequencies) and structural transformation (e.g., changes in semantic coupling), allowing political polarization to be examined not only as a shift in content or tone, but as a reorganization of the relationships that structure meaning within discourse.

2.6. *Data Preprocessing / Standardization*

Upon constructing each corpus, all transcripts were subjected to a standardized preprocessing and transcript cleaning pipeline to ensure comparability across both periods. A custom corpus-cleaning script for Python (see Appendix B) was used to remove non-linguistic artifacts, including timestamps, transcription notations, broadcast markers, and other platform-generated metadata. This step ensured that only linguistically meaningful content representing the video text remained for subsequent analysis. All preprocessing procedures were applied identically to both the 1980s and

2020s corpora, maintaining structural equivalence across datasets while reducing the risk that observed differences would be attributable to transcription conventions or formatting artifacts rather than substantive discourse change.

Following preprocessing, each corpus was uploaded to WMatrix7 (Rayson, 2008) for semantic annotation. Semantic tag frequencies were extracted across the four predefined semantic domains listed above (programmed into Wmatrix7), including: government, emotion, psychology, and social relations. These frequency distributions formed the basis for statistical comparison across time periods. Independent-samples t-tests were conducted using semtag frequencies between the two corpora to assess whether there were significant differences in the relative salience of each semantic domain. The use of t-tests also permitted comparison between mean frequency distributions across the two independent corpora, allowing for inference about whether observed differences were statistically reliable rather than due to sampling variation.

In addition to semantic tagging, a quantitative n-gram analysis was conducted for each corpus with a total of 1,000 four-gram sequences (4-grams) extracted per corpus, along with their associated frequency counts, Mutual Information (MI) scores, and log-likelihood (LL2) values. These measures were used to identify statistically salient collocational structures and recurrent lexical patterns within each period and to compare them diachronically. The resulting n-grams were then operationalized through a structured coding scheme (see Table 1), which mapped lexical patterns onto higher-order discourse categories. To ensure consistency and scalability in classification, a Python-based automation script (see Appendix C) was used to apply the coding scheme uniformly across both corpora.

Table 1. Coding Table for Functional Analysis of 4-gram Political Discourse.

Code	Category	Definition	Analytical Value
EPIS	Epistemic Stance	Expressions of certainty, uncertainty, or evaluation of knowledge	Captures speaker confidence, hedging, and rhetorical positioning
INTENT	Intentionality & Future Orientation	Statements expressing plans, commitments, or future actions	Reveals agency, commitment, and policy framing
INST	Institutional Reference	References to political institutions, roles, or formal entities	Indicates institutional grounding and formal political discourse
GEO	Geographic Reference	Mentions of locations or regions	Tracks spatial framing and regional political focus
PROC	Procedural / Framing Markers	Discourse markers structuring arguments or organizing speech	Reflects argumentative structure and rhetorical organization
QUANT	Quantification & Scale	Expressions indicating numerical or scalar magnitude	Shows shift from precise to approximate quantification
INTERACT	Interactional Language	Direct engagement with interlocutors or audience	Captures conversationalization and interpersonal tone

NEG	Negation & Opposition	Expressions of rejection, disagreement, or refusal	Indicates adversarial stance and political opposition
SOCREF	Social Reference	References to collective groups or populations	Reflects collective identity construction
ABILITY	Ability & Modality	Expressions of capacity or possibility	Captures modal reasoning and capability framing
DISFL	Disfluency / Spoken Features	Repetitions, contractions, or speech-like irregularities	Indicates shift toward informal, speech-like discourse
ECON	Economic / Policy Reference	References to financial or economic institutions or metrics	Tracks economic framing in political discourse

To move beyond frequency-based comparison and toward a relational model of discourse structure, the coded n-gram and semantic data were subsequently used to construct a network representation of discourse organization. In this model, nodes represent semantic domains and/or coded discourse categories derived from the WMatrix7 semantic fields and n-gram coding scheme. Edges represent statistically meaningful associations between these nodes, operationalized through Pearson correlation coefficients calculated on corpus-level semantic frequency distributions (e.g., semtag frequencies) across the four domains of government, emotional content, psychological processes, and social relations within each corpus. Pearson correlation captured the strength and direction of linear association between continuous frequency distributions, allowing for standardized comparison of coupling strength between semantic categories across time periods.

Coupling was operationalized as the degree of co-variation between two semantic and discourse categories across corpora: stronger positive correlations indicate tighter coupling, while weaker or negative correlations indicate greater semantic decoupling. With respect to their semantic networks, nodes were interpreted as corresponding to semantic or coded discourse categories, and edges to correlation-weighted relationships between these categories. This approach allowed discourse structures to be modeled as an interconnected semantic system, where shifts in edge strength and network configuration reflect broader diachronic changes in how political meaning is organized and distributed across semantic domains.

Following the evaluation of the semantic networks for each corpus, the analysis proceeded by comparing their structural properties to assess diachronic changes in discourse organization. Specifically, network-level metrics were computed to evaluate differences in overall connectivity, clustering, and centralization between the 1980s and 2020s corpora. Edge-weight distributions are examined to assess shifts in the strength and stability of semantic coupling, while community structure is analyzed to identify changes in the modular organization of discourse. Together, these measures allowed for the evaluation of whether contemporary political discourse exhibits greater fragmentation, reduced hierarchical organization, and weaker cross-domain integration relative to pre-digital political news discourse.

3. Results

Differences in semantic distributional patterns and network structure between the 1980s and 2020s political news corpora are reported below. Descriptive statistics (see Table 2) indicate broadly comparable distributional properties across the two corpora, with only a modest reduction in mean semantic frequency values in the 2020s dataset with minimal differences in variability, suggesting that any diachronic change is unlikely to be driven by simple shifts in overall word frequency distributions alone.

Table 2. Descriptive statistics comparison table.

	Frequency 1980	Frequency 2026
Mean	2.63	2.41
Mode	0.69	0.69
Std. Deviation	1.41	1.37
Minimum	0.69	0.69
Maximum	6.31	6.04
Quartile 1	1.61	1.39
Quartile 3	3.61	3.5
Number of values	127	127
95% Confidence interval for mean	2.38 - 2.87	2.17 - 2.65

Tests for normality indicated that the distribution for the 1980s sample (see Table 3 and Figures 2 & 3) deviated from normality, as evidenced by significant Shapiro-Wilk ($p < .001$) and Anderson-Darling ($p < .001$) tests, as well as the Lilliefors-corrected Kolmogorov-Smirnov test ($p = .002$), although the uncorrected Kolmogorov-Smirnov test was non-significant ($p = .127$). Tests for normality also indicated that distribution for the 2020s corpus deviated from normality (see Table 3 and Figures 4 & 5), as evidenced by significant Shapiro-Wilk ($p < .001$) and Anderson-Darling ($p < .001$) tests, as well as the Lilliefors-corrected Kolmogorov-Smirnov test ($p < .001$), although the uncorrected Kolmogorov-Smirnov test approached but did not reach significance ($p = .059$).

Taken together, these results indicate clear departures from normality in the 1980s and 2020s corpora. Accordingly, data were transformed using log transformation to reduce skewness and improve distributional symmetry, better satisfying assumptions for parametric statistical analyses. After transformation, variables (e.g., semtag frequencies for the corpora) exhibited mild positive skew for the 1980s corpus (skewness = 0.60) indicating mild to moderate right-skew and some asymmetry, and noticeable right-skew (still commonly manageable) for the 2020s corpus (skewness = 0.70), indicating that the frequencies for both are within tolerable limits for parametric analysis, particularly given the balanced sample sizes and the robustness of t-tests to moderate violations of normality.

Table 3. 1980 – Tests for Normal Distribution.

	Statistics	p
Kolmogorov-Smirnov	0.10	.127
Kolmogorov-Smirnov (Lilliefors Corr.)	0.10	.002

	Statistics	p
Shapiro-Wilk	0.95	<.001
Anderson-Darling	1.77	<.001

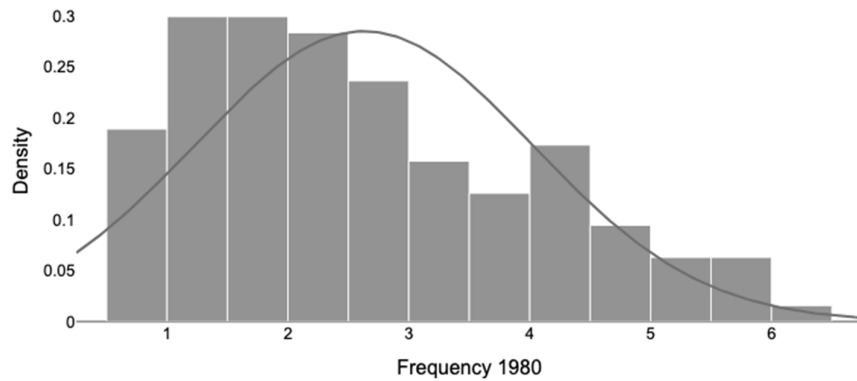


Figure 2. Histogram for Normal Distribution of Frequency of Semantic Features (1980s).

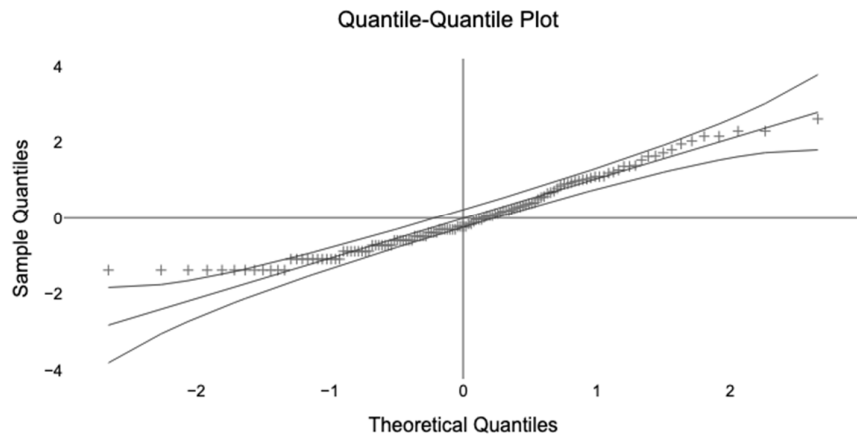


Figure 3. Quartile-Quantile (QQ) Plot Normal Distribution of Frequency of Semantic Features (1980).

Table 4. 2020s – Tests for Normal Distribution.

	Statistics	p
Kolmogorov-Smirnov	0.12	.059
Kolmogorov-Smirnov (Lilliefors Corr.)	0.12	<.001
Shapiro-Wilk	0.93	<.001
Anderson-Darling	2.56	<.001

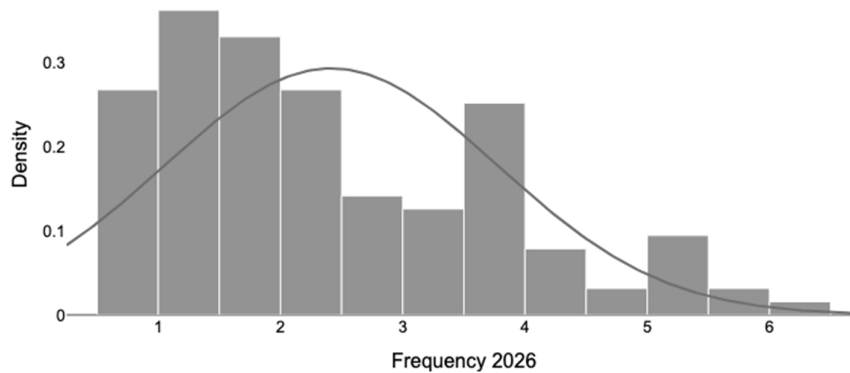


Figure 4. Histogram for Normal Distribution of Frequency of Semantic Features (2020s).

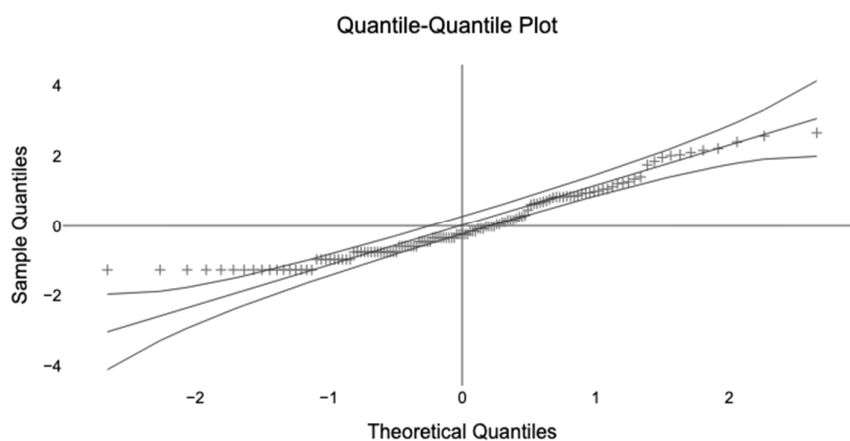


Figure 5. Quantile-Quantile (QQ) Plot Normal Distribution of Frequency of Semantic Features (2020).

3.1. Independent Sample T-Test

Tests for homogeneity of variance were conducted to assess whether the 1980s and 2020s corpora exhibited comparable variability in semantic frequency values, with results indicating no significant differences in dispersion between the two corpora. Levene's tests (see Table 4) based on the mean ($p = .702$) and median ($p = .669$), as well as the F-test ($p = .753$), all failed to reject the null hypothesis of equal variances. Together, these findings suggest that variability in semantic frequency distributions was statistically equivalent across the two time periods, despite minor differences in mean values and distributional shape.

The 1980s corpus had slightly higher scores ($M = 2.63$) than the 2020s corpus ($M = 2.41$), but this difference is small. The Levene's test showed that the assumption of equal variances was met (See Table 4). An independent samples t -test (see Table 5) found that the difference between the groups was not statistically significant ($p = .224$), meaning there is no reliable evidence of a real difference between them. The effect size ($d = 0.15$) is very small, indicating the difference is negligible in practical terms. Thus, there is no statistically significant difference between the semtag frequencies for the 1980s and the 2020s samples that can be explained by the frequencies of semtags alone.

Table 4. Levene test of variance equality.

Test	F	df1	df2	p
Levene's Test (Mean)	0.15	1	252	.702
Levene's Test (Median)	0.18	1	252	.669
F-Test	1.06	126	126	.753

The independent-samples t-test (see Table 5) likewise revealed no significant difference in mean semantic frequency values between the 1980s and 2020s corpora, $t(252) = 1.22$, $p = .224$, with a small effect size (Cohen's $d = 0.15$). The Welch-adjusted test produced identical results, confirming the robustness of the finding under unequal variance assumptions.

Table 5. t-Test for independent samples.

	t	df	p	Cohen's d
Equal variances	1.22	252.00	.224	0.15
Unequal variances	1.22	251.80	.224	0.15

The 95% confidence interval (see Table 6) for the mean difference between the 1980s and 2020s corpora ranged from -0.13 to 0.55, indicating that the interval included zero and therefore did not support a statistically reliable difference between groups. The observed mean difference was small ($M_{diff} = 0.21$), further confirming the absence of a meaningful divergence in semantic frequency values across time periods.

Table 6. 95% Confidence Interval of the Difference.

	Mean Difference	Standard Error of Difference	Lower limit	Upper limit
Equal variances	0.21	0.17	-0.13	0.55
Unequal variances	0.21	0.17	-0.13	0.55

Thus, given that t-test results were not significant, that the effect size was very small ($d = 0.15$), and that the Confidence Interval (CI) crosses zero, we can conclude that there is no meaningful distributional shift in semantic frequencies alone, suggesting, as anticipated, that any differences between the semantic quality of the 1980s and 2026 samples are likely relational and structural level (e.g., the level of semantic networks) rather than at the distributional level.

3.2. Pearson Correlation Analysis: Comparative Interpretation of Semantic Structure

Overall, the Pearson Correlation analysis across semantic fields (e.g., politics, emotions, psychology, social) for each corpus independently indicate a shift from a highly integrated semantic system in the 1980s toward a more fragmented and structurally differentiated system in 2020s (see Tables 7 & 8 for comparison), characterized by weaker coupling, increased variability, and the emergence of oppositional relationships between semantic domains (see Figures 6 and 7). In contrast to the 1980s corpus, which displays uniformly strong positive correlations indicative of a highly integrated semantic systems of meaning (see Table 8), the 2020s corpus demonstrates substantially greater variability in relational structure (see Table 10).

Specifically, for the 2020s corpus, strong positive coupling persisted between governance and social domains (GOVNT-SOC = .74), indicating continued alignment between institutional discourse and social framing at the collective level. Weak or near-zero relationships were observed between governance and emotional domains (GOVNT-EMO = .02), and between emotional and social domains, indicating increasing semantic independence between affective, institutional, and social semantic domains (see Figure 6). Negative correlations emerge between governance and psychological domains (GOVNT-PSYC = -0.37), and between psychological and social domains (PSYC-SOC = -0.22), suggesting antagonistic or inversely related semantic dynamics in 2026. Overall, these patterns suggest a shift from a highly integrated semantic networks of news media communication in the 1980s toward a more fragmented and structurally differentiated system in the 2020s, characterized by weaker coupling, increased variability, and the emergence of oppositional relationships between semantic domains.

Table 7. Correlation and significance across semantic domains for 1980s corpus

		GOVNT 1980	EMO 1980	PSYCH 1980	SOC 1980
GOVNT 1980	Correlation	1.00	0.94	0.98	0.97
	p-value		<.001	<.001	<.001
EMO 1980	Correlation		1.00	0.92	0.97
	p-value			<.001	<.001
PSYCH 1980	Correlation			1.00	0.94
	p-value				<.001
SOC 1980	Correlation				1.00
	p-value				

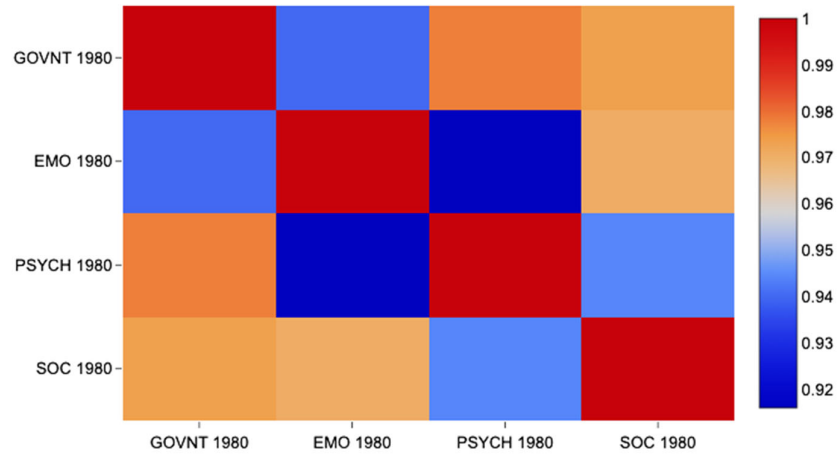


Figure 6. Correlation Heatmap for Semantic Domains in 1980s political news discourse.

Table 8. Correlation and significance across semantic domains for 2020s corpus.

		GOVNT 2026	EMO 2026	PSYC 2026	SOC 2026
GOVNT 2026	Correlation	1.00	0.02	-0.37	0.74
	p-value		.942	.173	.002
EMO 2026	Correlation		1.00	0.16	0.13
	p-value			.561	.651
PSYC 2026	Correlation			1.00	-0.22
	p-value				.427
SOC 2026	Correlation				1.00
	p-value				

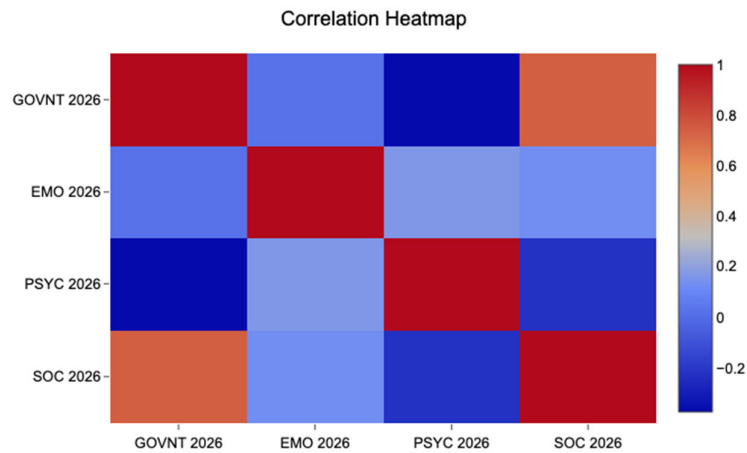


Figure 7. Correlation Heatmap for Semantic Domains in 2020s political news discourse.

3.3. N-gram Analysis Summary (1980 Corpus)

The 1980s corpus is most prevalently characterized by a highly formulaic and institutionally anchored discourse structure, reflected in strong mutual information (MI) values and distinctive log-likelihood (LL2) scores across recurrent 4-gram sequences (see Tables 9 & 10). Overall, patterns across n-gram frequencies, MI values, and LL2 scores indicate a highly routinized linguistic system organized around stable institutional references, structured epistemic positioning, and conventionalized rhetorical forms. Accordingly, the 1980s political news discourse exhibits high formulaicity and lexical stability, with frequent high-MI constructions such as “the house of commons,” “a great deal of,” and “it seems to me,” indicating reliance on entrenched, conventionalized phraseological units rather than spontaneous or emergent constructions. This suggests a strongly standardized discourse environment characterized by shared linguistic routines and predictable interactional alignment.

The 1980s corpus is likewise strongly anchored in institutional and geopolitical references, with high-frequency and highly distinctive phrases referencing formal political structures (e.g., “the house of commons,” “the bank of Canada,” “the liberal party,” “the province of Quebec”), which reflects a discourse system in which meaning is consistently grounded in stable institutional referents and national political frameworks. There is also evidence of collective and national-level framing, most clearly expressed through constructions such as “the people of Canada,” indicating a consistent rhetorical orientation toward an imagined collective public sphere in Canada rather than individualized or fragmented audiences. The discourse also demonstrates structured epistemic stance-taking, with recurrent hedging and evaluative constructions (e.g., “it seems to me,” “I don’t think,” “what I would do”), suggesting controlled argumentative positioning and moderated certainty rather than adversarial or emotionally reactive framing.

Table 9. 1980s n-gram analysis (4-gram).

N-gram	4 Words 1980	Freq	MI	LL2
1	of the liberal party	57	7.451	257.535
2	you said you would	97	5.052	195.135
3	you're going to	135	6.253	158.043
4	We're going to	98	6.253	158.043
5	the province of Quebec	78	7.743	143.405
6	the house of commons	35	10.292	142.675
7	it seems to me	21	8.496	139.534
8	the leader of the	52	4.487	138.69
9	I'd like to	68	7.686	133.522
10	what i would do	227	4.756	124.895
11	I would like to	227	5.677	111.431
12	in the province of	42	4.439	106.626

13	a great deal of	31	8.789	105.384
14	the bank of Canada	17	7.498	93.248
15	the people of Canada	129	5.115	83.98
16	on the one hand	135	7.567	59.867
17	Again we used to	15	4.603	49.181
18	I don't think	156	3.902	39.489
19	when it comes to	100	3.03	24.603
20	10 million a day	15	3.802	13.671

The 1980s corpus also displays deliberative and future-oriented agency, expressed through stable intentional constructions such as “we’re going to” and “I’d like to,” which reflect organized projections of action rather than fragmented or contested temporal orientations. Finally, the presence of logically structured argumentative scaffolding (e.g., “on the one hand,” “when it comes to”) and policy-anchored quantification (e.g., “10 million a day,” “a great deal of”) further indicate a discourse style organized around formal reasoning, institutional justification, and policy-oriented numerical framing. Taken together, the 1980s n-gram profile reflects a discourse system that is highly formulaic, institutionally embedded, collectively oriented, rhetorically structured, epistemically controlled, and policy/institution-focused. This corresponds to a highly integrated semantic-phraseological regime, in which linguistic patterns and meanings are stable, predictable, and strongly anchored in shared institutional and national frameworks.

Table 10. 2020s n-gram analysis (4-gram).

N-gram 4 Words 2026	Freq	MI	LL2
I don't	154	6.757	950.586
He's going to	127	4.724	538.228
we are going to	444	4.724	538.228
the honorable member for	17	10.608	229.635
Don't want an election	111	5.767	138.605
don't want to	154	5.767	135.895
they don't want	154	5.024	150.112
to be able to	17	5.377	132.542

I don't know	154	4.663	132.264
in terms of the	5	5.994	127.524
I don't think	17	5.03	113.882
There's a lot of	19	6.787	99.843
the course of the	17	5.469	90.374
the house of commons	5	10.558	73.184
over the course of	56	7.93	66.377
don't know if	108	4.175	55.337
of the steel that	16	4.556	39.709
that we want to	87	4.503	293.841
good to see you	51	5.4	27.986
We're going to have	127	3.184	21.477

A prominent feature of the 2020s corpus is the high density of negation-based constructions such as “don't want to,” “they don't want,” and “don't want an election.” These patterns reflect a discourse environment increasingly organized around oppositional and refusal-based framing, where political meaning is frequently articulated through rejection, disagreement, or constraint rather than structured argumentation or deliberation. Institutional references are present but less dominant and more contextually fragmented. While phrases such as “the house of commons” and “the honorable member for” remain highly distinctive, they are not embedded within a broader system of institutional phraseology as in the 1980s corpus. This suggests a partial decoupling of institutional language from wider discourse structures, where institutional references persist but are less systematically integrated into semantic organization.

The presence of conversational markers such as “good to see you” and informal discourse fragments indicates a shift toward interactional and phatic language use, suggesting that political discourse increasingly incorporates conversational and interpersonal registers rather than strictly formal or rhetorical structures. A notable cluster of constructions such as “over the course of,” “we are going to,” and “he's going to” reflects a strong orientation toward ongoing processes and projected action. However, unlike the structured future orientation in 1980s, these forms appear in a more fluid and less hierarchically organized discourse environment, often disconnected from stable institutional grounding.

Compared to the 1980s corpus, the 2020s n-gram analysis shows lower overall phraseological stability, with fewer tightly bound, high-MI institutional expressions and a greater proportion of flexible, context-dependent constructions (e.g., “there's a lot of,” “in terms of the,” “that we want to”). This indicates a shift toward less formulaic and more dynamically assembled discourse patterns. Overall, the 2020s n-gram profile reflects a discourse system that is more conversational and informal,

more epistemically uncertain and hedged, more negation and refusal-oriented, less institutionally integrated, more interactional and situationally responsive, with reduced concentration of high-MI institutional phraseology among dominant n-grams. In contrast to the highly structured and institutionally anchored 1980s discourse system, the 2020s corpus reflects a more heterogeneous and loosely coupled phraseological environment, where meaning is increasingly constructed through local interactional dynamics rather than stable institutional or rhetorical templates.

3.4. Corpus-Based Discourse Analysis

The 2020s corpus exhibits a marked redistribution of discourse functions relative to the 1980s corpus (see Tables 11 & 12). While institutional references remain stable in absolute frequency, they are no longer structurally dominant within the discourse system. Instead, the 2020s dataset is characterized by substantial increases in disfluency, negation, and conversational epistemic markers, indicating a shift toward more informal, fragmented, and interactionally oriented political language.

In contrast, the 1980s corpus demonstrates a more balanced and structurally integrated distribution of discourse functions, with comparatively lower levels of negation and disfluency, and stronger anchoring in collective, geographic, and institutional framing. The disappearance of social reference and geographic coding in the 2020s corpus further suggests a weakening of collective and spatial grounding in political discourse. Overall, these patterns indicate a transition from a relatively structured and institutionally embedded discourse system in the 1980s toward a more conversationally mediated, adversarial, and functionally dispersed system in the 2020s, consistent with increased semantic and pragmatic decoupling across domains. Thus, the 2020s corpus is not simply more informal or negative; it is structurally reorganized around negation, uncertainty, and interaction, rather than institutional reference, rhetorical structure, and collective anchoring.

Table 11. 1980s Code Frequencies.

ABILITY	1
ECON	1
EPIS	2
INST	6
INTENT	26
INTERACT	1
PROC	2
QUANT	1
DISFL	11
NEG	3
SOCREF	3
GEO	2

Table 12. 2020s Code Frequencies.

ABILITY	1
ECON	5
EPIS	2
INST	6
INTENT	36
INTERACT	1
PROC	4
QUANT	3
DISFL	57
NEG	32

Table 13 presents the distribution of coded functional categories based on the coding protocol outlined in Table 1. Because both corpora were sampled to equal size n-gram samples (n = 1000 4-grams each), raw frequencies are interpreted comparatively. Overall, these results indicate a clear diachronic shift in the functional composition of political discourse, with contemporary speech exhibiting increased emphasis on stance, interaction, and speech-like features, relative to the more institutionally grounded and structurally formal discourse observed in the 1980s corpus (see Tables 13 & 14 for comparison). Some discourse features remain stable across time. Institutional references (INST), for instance, occur at identical frequencies in both corpora (6 to 6), indicating that formal political entities (e.g., parliamentary bodies and roles) continue to anchor political discourse. Similarly, epistemic stance (EPIS) markers show no change (2 to 2), suggesting that expressions of uncertainty or evaluation persist but do not increase in frequency.

In contrast, several categories exhibit substantial change. Intentionality and future-oriented expressions (INTENT) increased from 26 in the 1980s to 36 in the 2020s, reflecting a stronger emphasis on planned action, commitment, and forward projection in contemporary discourse. This pattern aligns with the increased prevalence of constructions such as “we are going to” and “they don’t want,” which foreground agency and anticipated outcomes. One of the most pronounced shifts is observed in negation (NEG), which increases more than tenfold (3 in 1980s to 32 in the 2020s), suggesting a significant rise in oppositional and adversarial framing in contemporary political language. Rather than framing arguments through structured contrast (e.g., “on the one hand”), speakers in the 2020s corpus more frequently employ direct rejection and refusal (e.g., “don’t want,” “do not think”).

This shift reflects a movement away from formally balanced argumentation toward more immediate, stance-driven disagreement. The most dramatic change occurred in disfluency and spoken-language features (DISFL), which increase from 11 in the 1980s to 57 in the 2020s. These include repetitions, fragmented constructions, and contraction artifacts (e.g., “I, I don’t,” “he’s going to”), all of which are characteristic of spontaneous or speech-like production. This finding provides strong support for the conversationalization of political discourse, whereby contemporary political speech increasingly resembles informal spoken or digitally mediated interaction rather than carefully structured institutional rhetoric. The co-occurrence of disfluency with increased interactional markers (e.g., INTERACT: 1 to 1, but qualitatively more conversational forms) further supports this interpretation.

The 1980s corpus also contains categories that are entirely absent in the 2020s, notably geographic reference (GEO: 2 to 0) and social reference (SOCREF: 3 to 0). This suggests a decline in explicit references to collective identity (e.g., “the people of Canada”) and regional framing (e.g., “the

province of Quebec”), which also suggests a reduced emphasis on spatial and collective anchoring. At the same time, procedural markers (PROC) increase modestly (2 to 4), though their form changes qualitatively. Earlier structured discourse markers (e.g., “on the one hand”) are replaced by looser framing devices (e.g., “in terms of the,” “over the course of”), suggesting a shift toward less rigid argumentative organization.

Additional shifts are observed in categories related to abstraction and policy framing. Quantification (QUANT) increases (1 to 3), but shifts from precise numerical expressions (e.g., “10 million a day”) to more approximate forms (e.g., “there’s a lot of”), indicating reduced numerical specificity. Economic references (ECON) also increase (1 to 5), suggesting greater integration of economic framing in contemporary discourse. Meanwhile, ability and modality (ABILITY) emerge only in the 2020s corpus (0 to 1), reflecting increased use of constructions such as “to be able to,” which emphasize capability and feasibility.

Taken together, these results point to a systematic transformation in the functional profile of political news discourse. The 1980s corpus is characterized by institutional and geographic grounding, structured rhetorical organization, and collective and referential framing. In contrast, the 2020s corpus exhibits increased intentionality and forward projection, strongly elevated negation and adversarial stance, high levels of disfluency and speech-like fragmentation, and reduced reliance on collective and spatial references. These patterns suggest a broader shift from formal, institutionally anchored discourse toward a more interactional, performative, and conversational modes of political communication, in which speakers prioritize stance expression, immediacy, and audience engagement over structural formality.

Most importantly, these findings converge across analytical levels. While distributional analyses show no significant differences in semantic frequency, network analysis reveals a collapse in cross-domain coupling, and n-gram analysis demonstrates a shift toward interactional, negation-driven, and disfluent language. Taken together, these results suggest that diachronic change in political discourse is not primarily lexical or quantitative, but structural, manifesting as a reorganization of how meaning is constructed across semantic domains.

Discussion

This study examined whether the semantic dynamics of contemporary political news discourse reflect greater fragmentation and identity dislocation in political news discourse relative to pre-digital, broadcast-era media environments. Findings consistently support this thesis across all levels of analysis. While the distribution of semantic content remains stable over time, the relational organization of meaning has undergone substantial transformation.

At the micro-level, the 2020s corpus exhibits significantly weaker coupling across semantic domains compared to the 1980s, alongside greater instability in cross-domain relationships. Whereas the 1980s corpus is characterized by tightly integrated and uniformly correlated structures, the 2020s data reveal attenuated and, at times, oppositional relationships between domains. This pattern indicates reduced category coherence, increased boundary permeability, and heightened semantic volatility, consistent with more fragmented meaning structures. At the meso-level, these dynamics correspond to patterns of semantic contagion and identity diffusion. In the 2020s corpus, identity-related terms are more weakly coupled and less consistently co-articulated across domains (a trend also noted in Gündüz, 2017 [29], and Jordan, 2020 [30]). Rather than reinforcing stable group identities, meanings circulate without durable integration, producing identity representations that are more fluid, context-dependent, and structurally diffuse (which aligns with findings in Jordan, 2020 [31], and Rehman, 2025 [32]).

At the system level, weakened semantic coupling is associated with a reorganization of normative structures. Although normative content remains present, it appears denser but less hierarchically organized, reflecting diminished centrality of shared evaluative anchors and the coexistence of competing or loosely connected normative frameworks. This pattern is consistent with increased normative ambiguity, where expectations persist but lack clear structural coordination. At

the macro-level, the relationship between institutional actors and legitimacy-related constructs is markedly weaker in the 2020s corpus. In contrast to the 1980s, where institutional discourse is embedded within coherent social and normative frameworks, contemporary discourse shows reduced alignment between institutions and broader meaning structures. This suggests a form of erosion in coherence across semantic domains, in which the relationships between institutions and legitimacy are no longer consistently maintained within the public sphere.

Taken together, these findings reveal a clear divergence between distributional stability and structural change. Across corpora, there is no significant difference in content and frequency, indicating continuity at the level of substance. However, this stability conceals a fundamental reorganization in how meaning is structured. The 1980s discourse reflects a highly integrated system in which institutional, emotional, psychological, and social dimensions cohere within a unified interpretive framework, consistent with the centralized logic of broadcast media.

By contrast, the 2020s corpus exhibits a decoupled and modular configuration. Semantic domains no longer align systematically but operate more independently, producing localized and context-dependent associations rather than a globally coherent structure. This shift aligns with the transformation of the contemporary media environment. Political discourse now unfolds within algorithmically mediated systems that curate, personalize, and dynamically redistribute political information across networked publics. These systems restructure the conditions under which meaning is produced, fragmenting communicative space into partially disconnected semantic environments. Thus, the co-articulation of institutional, social, emotional, and psychological meanings becomes attenuated, producing the weakened correlations observed in the data.

These structural changes carry several implications. Political discourse becomes more functionally differentiated, with domains serving specialized and less coordinated roles in the contemporary context. Weak and negative correlations suggest that domains may compete rather than reinforce one another, reflecting increased personalization and oppositional framing. Coherence persists, but only locally within selective domain relationships rather than across the semantic system as a whole. Importantly, these transformations occur without any change in overall frequency of content itself, indicating that shifts in political discourse are not primarily lexical or affective, but relational.

In this sense, contemporary polarization is better understood not simply as ideological divergence, but as a breakdown in how we structurally integrate signifiers (e.g., words and phrases) to socially construct political meaning using language within the public sphere. The observed semantic decoupling provides empirical grounding for a condition in which shared interpretive frameworks weaken, and the coherence necessary for collective understanding and democratic coordination is progressively undermined, not by the kind of content that is circulated, but by the ways in which that content is assembled into meaningful wholes.

Theoretical Implications

With respect to theoretical development, the findings outlined above also suggest that contemporary political discourse is organized less as a unified semantic field and more as a loosely coupled system of semi-autonomous components. The transition from coherence to fragmentation reflects a move away from integrated meaning-making toward a more differentiated, modular, and sometimes antagonistic structure of discourse. Within the framework of this study, this can be understood as a shift from semantic coupling toward increased semantic decoupling, from discursive integration toward discursive fragmentation, and from institutional coherence to more functional specialization across parties in more antagonistic fashion.

Conclusions

The analysis and findings presented across this paper are consistent with the broader transformations noted in existing research on political communicative environments (e.g., Ferrara et

al., 2016 [15]; Stella et al., 2018 [16]; Mansoury et al., 2020 [17]; Woolley, 2020 [18]; Prior, 2007 [19]; Bennett & Iyenga, 2008 [20]; Klinger & Svensson, 2015 [21]; Van Aelst et al., 2017 [22]; Van Dijck, Poell & de Waal, 2018 [23]). This study builds on this research by providing additional substantiation for the claim that contemporary political discourse is characterized by increased interactivity, fragmentation of audiences, and the coexistence of multiple discourse styles within news media ecosystem in the digital era. Findings presented in this paper demonstrate that political discourse has remained quantitatively stable over time but has undergone a qualitative transformation in its semantic structure and organization across political communication networks, characterized by a shift from coherent and integrated semantic systems to fragmented, modular, and differentiated systems of meaning. This distinction between surface stability and structural change is the main contribution that this paper makes, which fundamentally demonstrates that meaningful diachronic transformation in political news discourse may not be visible at the level of frequency of the content alone, but emerges clearly when examining the relationships that organize meaning across domains.

Concomitantly, implications of this transformation in the semantic networks of political discourse extend well beyond the problem of polarization as ideological divergence, pointing instead to a more fundamental reconfiguration of the conditions under which shared meaning is produced and reproduced in the digitized public sphere. In the broadcast-era model (1980s), political discourse operated within a relatively integrated semantic system, allowing disagreement to unfold within common interpretive frameworks. By contrast, the contemporary communicative ecosystem is characterized by fragmented and weakly coupled semantic structures, in which meanings circulate without consistent integration or anchoring across domains. Under these conditions, political conflict is less moderated by shared points of reference, making disagreement more difficult to resolve and collective understanding more difficult to sustain.

At the broader level of collective identity, this shift has significant consequences for both collective consciousness and collaborative cognition within the public sphere, as well as democratic and social cohesion more generally. The weakening of semantic integration may undermine the coherence of shared normative frameworks, leading to the destabilization of the relationship between institutions and legitimacy while contributing to more diffuse and context-dependent constructions of national and regional identities. In this sense, this paper points how semantic networks of digitally mediated political discourse shapes socio-political culture and conditions that are analogous to anomie, not as an absence of meaning, but as a breakdown in the structural integration of meaning across social domains. This means that semantic cohesion becomes localized and nuclear rather than systemic, thus rendering our collective capacity for democratic coordination and cooperation across partisan divides increasingly constrained by the fragmentation of the semantic infrastructure through which collective life is organized.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A: Transcript Sources for Comparative Analysis of CBC News Coverage (1980s vs. 2020s)

Transcript Source 1980s	Transcript Source 2020s
CBC News. 1984, February. <i>PET 1984</i> . [Video]. https://www.youtube.com/watch?v=e2QMS7SvAZ4	CBC. 2025, April 28 <i>Liberals projected to win 4th term, but unclear if minority or majority</i> <i>Canada Votes 2025</i> [Video]. https://www.youtube.com/watch?v=SATBOqyYODU

- CBC News. 1979, December 13. *When a federal budget in 1979 triggered an election* [Video].
<https://www.youtube.com/shorts/q6xh7ht3s7M>
- CBC News. 1979, May 13. *Canada Vote 1979* [Video].
<https://www.youtube.com/watch?v=j9dwE15I260>
- CBC News. 1988. *Betting On Free Trade 1988* [Video].
<https://www.youtube.com/watch?v=gyYjRmM7RDY>
- CBC News. 1983, June 11. *Mulroney & Clark 1983* [Video].
<https://www.youtube.com/watch?v=Uf90BCM7sbc>
- CBC News. 1981. *Constitutional Criticism 1981* [Video].
<https://www.youtube.com/watch?v=HLaGEHrDWx0>
- CBC News. 1983, April. *Pierre Trudeau and the media boycott, 1983* [Video].
<https://www.youtube.com/watch?v=AgmALhdyLYI&t=4s>
- CBC News. 1981, November 5
Pierre Trudeau gives an update on the Canadian Charter of Rights and Freedoms [Video].
<https://www.youtube.com/watch?v=8nInSdlteMk&t=33s>
- CBC The Natational. Feb. 07, 1989. *The Journal* [Video].
<https://www.youtube.com/watch?v=ukogTtmdA60>
- CBC Power & Politics. (2025, April 22). How do the Liberal and Conservative platforms compare? [Video]. YouTube.
<https://www.youtube.com/watch?v=DGBgmov4b60>
- CBC News, The National. (2025, October 8). Breaking down Alberta's use of the notwithstanding clause [Video]. YouTube.
https://www.youtube.com/watch?v=y9MGDpaK_2o
- CBC News. (2021, March 22). O'Toole promises 'comprehensive' climate plan 'before an election' [Video]. YouTube.
https://www.youtube.com/watch?v=_T_FeWWXoXs
- CBC News Special. (2025, November 4). Federal Budget 2025 [Video]. YouTube.
<https://www.youtube.com/watch?v=ICAvy71kGP8>
- CBC Power & Politics. (2026, February 27). Canada's economy contracted in the fourth quarter of 2025 [Video]. YouTube.
https://www.youtube.com/watch?v=JjFHIL_4T5w
- CBC News. (2025, September 10). Carney says global economy going through a 'rupture' [Video]. YouTube.
<https://www.youtube.com/watch?v=Fgz4nKrAjz4>
- CBC Power & Politics. (2025, September 10). Poilievre courts delegates as he faces a must-win leadership review vote [Video]. YouTube.
<https://www.youtube.com/watch?v=Do1HAojuw80&t=1s>

Appendix B: Python Script 1 Python Script for Text Preprocessing and Corpus Cleaning

```
import re

# =====
# 1. PASTE YOUR RAW TEXT HERE
```

```

# =====
raw_text = ""

"""

# =====
# 2. CLEANING FUNCTIONS
# =====

def remove_urls(text):
    return re.sub(r'http\S+|www\S+', "", text)

def remove_mentions(text):
    return re.sub(r'@\w+', "", text)

def remove_hashtags(text, keep_text=True):
    if keep_text:
        # Keep the word but remove #
        return re.sub(r'#(\w+)', r'\1', text)
    else:
        return re.sub(r'#\w+', "", text)

def remove_emojis(text):
    emoji_pattern = re.compile(
        "[
        \U0001F600-\U0001F64F" # emoticons
        \U0001F300-\U0001F5FF" # symbols & pictographs
        \U0001F680-\U0001F6FF" # transport & map
        \U0001F1E0-\U0001F1FF" # flags
        ]+",
        flags=re.UNICODE
    )
    return emoji_pattern.sub("", text)

def remove_extra_whitespace(text):
    return re.sub(r'\s+', ' ', text).strip()

def remove_punctuation(text):
    return re.sub(r'[^\w\s]', "", text)

# =====
# 3. APPLY CLEANING PIPELINE
# =====

cleaned_text = raw_text
cleaned_text = remove_urls(cleaned_text)
cleaned_text = remove_mentions(cleaned_text)
cleaned_text = remove_hashtags(cleaned_text, keep_text=True)
cleaned_text = remove_emojis(cleaned_text)
cleaned_text = remove_punctuation(cleaned_text)
cleaned_text = remove_extra_whitespace(cleaned_text)

# =====

```

```
# 4. SAVE TO .TXT FILE
# =====

output_file = "cleaned_corpus.txt"

with open(output_file, "w", encoding="utf-8") as f:
    f.write(cleaned_text)

print(f"Cleaned text saved to {output_file}")
```

Appendix B: Python Script 2 for N-Gram Coding and Analysis

```
import pandas as pd
import re
# -----
# 1. LOAD DATA
# -----
file_path = "/Users/sophiaricciardone/Documents/Publications/Information/Data/ngrams.csv"
df = pd.read_csv(file_path)

# Inspect and standardize column names
print("Original columns:", df.columns.tolist())
df.columns = df.columns.str.strip().str.lower()
print("Cleaned columns:", df.columns.tolist())

# Ensure required column exists
if "ngram" not in df.columns:
    raise ValueError("Column 'ngram' not found. Check CSV header formatting.")

# Normalize text
df["ngram_clean"] = df["ngram"].astype(str).str.lower().str.strip()

# -----
# 2. DEFINE CODING RULES
# -----

coding_rules = {
    "EPIS": [
        r"\bi do not know\b", r"\bi do nt know\b",
        r"\bi do not think\b", r"\bi do nt think\b",
        r"\bit seems to me\b", r"\bwhat i would do\b"
    ],
    "INTENT": [
        r"going to", r"would like to",
        r"we are going to", r"we're going to", r"he s going to"
    ],
    "INST": [
        r"house of commons", r"bank of canada",
        r"honorable member", r"leader of the"
    ],
    "GEO": [
        r"province of quebec", r"in the province of"
    ],
}
```

```

"PROC": [
    r"on the one hand", r"when it comes to",
    r"in terms of the", r"over the course of", r"the course of the"
],
"QUANT": [
    r"\d+ million", r"a great deal of",
    r"there s a lot of", r"theres a lot of"
],
"INTERACT": [
    r"you said you would", r"good to see you"
],
"NEG": [
    r"do nt", r"don't", r"do not",
    r"nt want", r"do nt want"
],
"SOCREF": [
    r"people of canada"
],
"ABILITY": [
    r"to be able to"
],
"DISFL": [
    r"\bi i\b", r"\bhe s\b", r"\bwere\b", r"\bnt\b"
],
"ECON": [
    r"bank of canada", r"steel"
]
}

```

```

# -----
# 3. CODING FUNCTION
# -----

```

```

def assign_codes(text):
    """
    Assigns one or more discourse codes to an n-gram
    based on predefined regular expression patterns.
    """
    codes = []
    for code, patterns in coding_rules.items():
        for pattern in patterns:
            if re.search(pattern, text):
                codes.append(code)
                break
    return list(set(codes))

```

```

# Apply coding
df["codes"] = df["ngram_clean"].apply(assign_codes)

```

```

# -----
# 4. EXPAND MULTIPLE CODES

```

```

# -----

df_expanded = df.explode("codes")

# -----
# 5. SUMMARY OUTPUT
# -----

code_counts = df_expanded["codes"].value_counts()

if "year" in df.columns:
    code_year = (
        df_expanded
        .groupby(["year", "codes"])
        .size()
        .unstack(fill_value=0)
    )
else:
    print("Warning: 'year' column not found — skipping year-based analysis.")
    code_year = None
# -----
# 6. SAVE OUTPUTS
# -----

df.to_csv("coded_ngrams_full.csv", index=False)
df_expanded.to_csv("coded_ngrams_expanded.csv", index=False)
code_counts.to_csv("code_frequencies.csv")
if code_year is not None:
    code_year.to_csv("code_by_year.csv")

# -----
# 7. PRINT SUMMARY
# -----

print("\nCode Frequencies:\n")
print(code_counts)

if code_year is not None:
    print("\nCode Distribution by Year:\n")
    print(code_year)

```

References

1. Pew Research Center (2014). Political Polarization in the American Public. <https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/>
2. Durkheim, É. (2002). *Suicide: A study in sociology* (J. A. Spaulding & G. Simpson, Trans.). Routledge. (Original work published 1897)
3. Merton, R. K. (1938). Social structure and anomie. *American Sociological Review*, 3(5), 672–682.
4. Gündüz, U. (2017). The Effect of Social Media on Identity Construction. *Mediterranean Journal of Social Sciences*, 8(5), 85–92. <https://doi.org/10.1515/mjss-2017-0026>
5. Jordan, K. (2020). Imagined audiences, acceptable identity fragments and merging the personal and professional: How academic online identity is expressed through different social media platforms. *Learning, Media and Technology*, 45(2), 165–178. <https://doi.org/10.1080/17439884.2020.1707222>

6. Rehman, A. (2025). Social Media and Youth Identity Formation. *Journal of Language, Literature & Social Affairs*, 1(1), 10–19. <https://doi.org/10.63056/jllsa.1.1.2025.4>
7. Abercrombie, G., & Batista-Navarro, R. (2020). Sentiment and position-taking analysis of parliamentary debates: A systematic literature review. *Journal of Computational Social Science*, 3(1), 245–270. <https://doi.org/10.1007/s42001-019-00060-w>
8. Németh, R. (2023). A scoping review on the use of natural language processing in research on political polarization: Trends and research prospects. *Journal of Computational Social Science*, 6(1), 289–313. <https://doi.org/10.1007/s42001-022-00196-2>
9. Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & Quantity*, 51(6), 2623–2646. <https://doi.org/10.1007/s11135-016-0412-4>
10. Farzindar, A. A., & Inkpen, D. (2020). *Natural Language Processing for Social Media*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-02175-6>
11. Camacho-collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., Boisson, J., Espinosa Anke, L., Liu, F., & Martínez Cámara, E. (2022). TweetNLP: Cutting-Edge Natural Language Processing for Social Media. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–49. <https://doi.org/10.18653/v1/2022.emnlp-demos.5>
12. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
13. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
14. Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, 27(2), 212–225. <https://doi.org/10.1017/S0140525X04450055>
15. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>
16. Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. <https://doi.org/10.48550/ARXIV.1802.07292>
17. Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, R. (2020). Feedback Loop and Bias Amplification in Recommender Systems. *arXiv*. <https://doi.org/10.48550/ARXIV.2007.13019>
18. Woolley, S. C. (2020). Bots and Computational Propaganda: Automation for Communication and Control. In N. Persily & J. A. Tucker (Eds.), *Social Media and Democracy* (1st ed., pp. 89–110). Cambridge University Press. <https://doi.org/10.1017/9781108890960.006>
19. Prior (2007)
20. Bennett, W. L., & Iyengar, S. (2008). A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication*, 58(4), 707–731. <https://doi.org/10.1111/j.1460-2466.2008.00410.x>
21. Klinger, U., & Svensson, J. (2015). The emergence of network media logic in political communication: A theoretical approach. *New Media & Society*, 17(8), 1241–1257. <https://doi.org/10.1177/1461444814522952>
22. Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., De Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stepińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheaffer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, 41(1), 3–27. <https://doi.org/10.1080/23808985.2017.1288551>
23. Dijk, J. van, Poell, T., & Waal, M. de. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.
24. Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424. <https://doi.org/10.1017/S1930297500005271>
25. Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
26. Bottou, L., Curtis, F. E., & Nocedal, J. (2016). Optimization Methods for Large-Scale Machine Learning. *arXiv*. <https://doi.org/10.48550/ARXIV.1606.04838>

27. Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. <https://doi.org/10.48550/ARXIV.1404.7828>
28. Mason, L. (2018). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.
29. Hedayatifar, L., Rigg, R. A., Bar-Yam, Y., & Morales, A. J. (2019). U.S. Social Fragmentation at Multiple Scales (arXiv:1809.07676). arXiv. <https://doi.org/10.48550/arXiv.1809.07676>
30. Kossowska, M., Klodkowski, P., Siewierska-Chmaj, A., Guinote, A., Kessels, U., Moyano, M., & Strömbäck, J. (2023). Internet-based micro-identities as a driver of societal disintegration. *Humanities and Social Sciences Communications*, 10(1), 955.
31. Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549. <https://doi.org/10.1075/ijcl.13.4.06ray>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.