Exploring COVID-19 Framing Across Diverse Platforms: Analyzing Semantic and Contextual Shifts in Public Discussion, News Media, and Government Communication

Hanjing Shi[1], Zhila Aghajari[1], Dominic DiFranzo[1], Haiyan Jia[1], Eric P. S. Baumer[2*]
[1]Lehigh University
[2]Faculty of Information, University of Toronto
{hasa23, zha219, djd219, haj616}@lehigh.edu
eps.baumer@utoronto.ca

Abstract

Understanding how shared issues are framed differently across public, news, and government discourse is central to the study of COVID-19 communication. This paper uses the previously validated Linked Latent Theta Role (LLTR) model as part of a novel analytic technique to examine framing differences across Reddit posts, mainstream news articles, and state public health bulletins. Rather than introducing a new model, we operationalize LLTR outputs to compare cross-source framing by examining how shared topic words are embedded in different syntactic constructions. Using a source-balanced corpus, we measure cross-source differences using Jensen–Shannon divergence over distributions of dependency-based relation–argument pairs, and contrast these results with a lexical baseline. Across eight COVID-19 topics, we find that sources often rely on overlapping topic vocabularies, yet diverge substantially in their syntactic realizations of those topics. Inspection of high-divergence grammatical evidence reveals systematic differences in how sources assign agency, attribute responsibility, and structure evaluative context around shared topical concepts. These findings suggest that grammar-aware representations provide an interpretable and scalable basis for identifying framing differences that are not visible at the lexical level alone.

Introduction

The COVID-19 pandemic generated an extraordinary volume of public discourse across social media, mainstream journalism, and official government communications. For example, more than 100 million COVID-related tweets were posted in the first few months of 2020 alone (Chen, Lerman, and Ferrara 2020). Each platform addresses a different audience with distinct communicative goals, leading to divergent linguistic styles and framing practices. Public health agencies tend to emphasize factual guidance and risk communication, whereas news outlets focus on reporting developments and amplifying salient events (Song et al. 2025). As communication scholars have noted, how public health issues are framed can shape public interpretation, even when the underlying information remains similar (Ogbodo et al. 2020).

While prior research on COVID-19 communication has produced many insights, it has typically examined each platform or modality in isolation (Tahamtan et al. 2021). Numerous content analyses have documented how the news media framed the pandemic, often emphasizing crisis, fear, and consequences (Ogbodo et al. 2020; Ebrahim 2022), and some studies have identified prevalent frames in social media discourse (Tahamtan et al. 2021; Sussman et al. 2023; Bergenfeld 2024). However, far fewer works have systematically compared framing across the broader media ecosystem. Tahamtan et al. observe that past studies concentrated on traditional news outlets and may not capture the diverse perspectives voiced on social networks (Tahamtan et al. 2021). This leaves a gap in understanding **how the same topics (e.g., masks, vaccines, case data) are portrayed differently** by, for example, a Reddit user, a news article, or a government press release. Examining such cross-platform framing in retrospect is essential for capturing how shared topic terms acquire different meanings over time.

To address this gap, the present study analyzes differences in framing of shared topics and keywords across platforms. By "framing," we adopt Entman's classic definition: "to select some aspects of a perceived reality and make them more salient in a communicating text" (Entman 1993). In practice, framing involves not only selecting which topics to discuss, but also how those topics are described, e.g., who is implied to be acting (or acted upon). We extend the linguistic perspective of Baumer et al., who demonstrated that semantic cues (e.g. factive verbs, hedging, entailments) can help classify framing in texts (Baumer et al. 2015). In our view, grammatical and semantic structures are highly indicative of framing—the choice of subjects, objects, modifiers, and verbs that construct narratives about actions and outcomes. For example, a previous study found that war metaphors were extensively used by non-experts on Twitter to describe COVID-19 (e.g. 'fighting an invisible enemy'), highlighting a combative frame of the pandemic (Wicke and Bolog-

*Eric P. S. Baumer was affiliated with Lehigh University when this work was completed and with the University of Toronto when it was published.

nesi 2020). In contrast, official health communications often avoided militaristic terms, favoring a more instructional or community-oriented tone (Kandzer et al. 2022). Such variations suggest that the same keyword—say, “mask” or “vaccine”—could be framed as an act of personal responsibility on one platform, but as a politicized symbol or a public mandate on another.

Methodologically, this analysis builds on the Linked Latent Theta Role (LLTR) model (Aghajari 2025, Ch. 4), which was developed to capture syntactic regularities in topic-specific language at scale. LLTR is neither introduced nor modified in this paper. Instead, it serves as a syntactic lens for our analysis, allowing us to examine how grammatical roles and their associated arguments vary across sources and can be used as systematic evidence for comparing framing across heterogeneous communication sources.

The model jointly learns topic–word distributions and the grammatical contexts in which those words appear, including the roles they occupy (e.g., subject, object, modifier) and the arguments they are linked to. For example, a term such as “coronavirus” may function as a nominal modifier in institutional reporting (e.g., “coronavirus pandemic”), while appearing as an active participant in experiential or causal constructions in other contexts (e.g., something that “spreads” or “infects” people). These differences reflect distinct ways of assigning agency, responsibility, and urgency to—that is, distinct ways of framing—the same underlying events.

Using LLTR, we identify topic keywords that recur across Reddit posts, mainstream news articles, and government texts, and examine how their grammatical evidence differs by source. To enhance interpretability, we introduce a source-aware filtering procedure based on normalized argument frequency, which retains only those role–argument configurations that are unusually salient within a given source. The resulting distributions capture characteristic framing tendencies for each communication environment.

We then compare these source-specific framing patterns using Jensen–Shannon divergence over normalized grammatical evidence distributions. This computational analysis is complemented by qualitative inspection of representative examples, allowing us to interpret how specific grammatical choices contribute to framing differences. This approach connects large-scale syntactic patterns with their linguistic instantiations, supporting systematic comparisons of how COVID-19 narratives are constructed and differentiated across public, media, and governmental discourse.

This analysis addresses two primary research questions concerning cross-platform variation in framing and its systematic identification through linguistic structure.

1. How is the COVID-19 pandemic framed differently across Reddit, news media, and government communications?

2. How do grammatical roles, argument structures, and topical emphases indicate differences in framing across communicative contexts?

Literature Review

Media Framing of COVID-19 Across Platforms

Framing theory emphasizes that how issues are characterized by communicators shapes public interpretation by highlighting particular problem definitions, causal attributions, and evaluative orientations (Hallahan 1999; Scheufele 1999; Chong and Druckman 2007). During the COVID-19 pandemic, a substantial body of research has examined how different communication arenas framed the crisis, revealing systematic differences across news media, social media, and official government communication.

Mainstream news coverage largely framed COVID-19 through crisis- and consequence-oriented narratives, emphasizing fear, uncertainty, and human impact. Content analyses of early pandemic coverage show the dominance of fear-related and human-interest frames, with limited use of reassuring or recovery-oriented framing (Ogbodo et al. 2020; Ebrahim 2022; Rooke 2021). These patterns are consistent with prior research on epidemic reporting, in which professional journalism foregrounds urgency, conflict, and societal disruption.

Social media discourse, by contrast, exhibited more heterogeneous and vernacular framing. Studies of platforms such as Twitter document pervasive metaphorical framing, notably war metaphors that conceptualize the virus as an enemy to be fought, as well as emotionally charged and evaluative narratives that often diverge from official public health messaging (Wicke and Bolognesi 2020; Panzeri, Di Paola, and Domaneschi 2021; Tsao et al. 2021). Such work highlights the role of social platforms in amplifying experiential, affective, and sometimes polarizing framings.

Official government and public health communication generally emphasized guidance, collective responsibility, and compliance-oriented framing. Analyses of agency messaging during the pandemic show a predominance of instructive, empathetic, and action-oriented frames designed to sustain public trust and encourage protective behavior (Kandzer et al. 2022; Rao et al. 2020). Comparative studies further indicate that even when operating on the same platforms, public health agencies tend to adopt more policy- and guidance-focused framing than news organizations, which emphasize case updates and broader pandemic impacts (Song et al. 2025).

Cross-Platform Comparisons and Methodological Approaches

While early COVID-19 framing studies typically examined individual communication arenas in isolation, subsequent work has increasingly compared narratives across platforms and actors. Qualitative analyses have shown that governments, news media, and other actors construct distinct narrative roles and storylines around the pandemic, but such approaches are limited in scale and do not quantify the prevalence of framing patterns (Holmes 2025; Mohammadi et al. 2022). Quantitative content analyses scale comparison by coding predefined frames across sources, revealing systematic differences between political, media, and institutional actors, but require extensive manual annotation and rely on fixed frame taxonomies (Ogbodo et al. 2020; Nienhaus 2024).

Computational approaches further scale cross-platform analysis using topic modeling, correspondence analysis, or attention-based measures. These studies show that public, media, and institutional agendas often diverge, even when addressing the same events, but typically infer framing indirectly from topics, keywords, or attention dynamics (Han, Yang, and Piterou 2021; Cinelli et al. 2020; Reveilhac 2022). As a result, they capture broad thematic differences while offering limited access to the linguistic mechanisms through which framing is enacted (Chong and Druckman 2007; Nicholls and Culpepper 2021).

A smaller set of mixed-method studies combines automated analysis with manual interpretation to balance scale and interpretability. These works indicate that computational models capture dominant frames but often miss subtler or less frequent framing patterns, motivating approaches that better connect quantitative outputs to qualitative interpretation (Hayek 2024; Kermani et al. 2024). Recent surveys of computational framing analysis further note that most large-scale approaches rely on lexical, semantic, or topic-level features, with limited attention to grammatical structure as a locus of framing variation (Ali and Hassan 2022; Nicholls and Culpepper 2021).

Positioning This Study

Building on this literature, our study addresses two key gaps. First, we provide a systematic cross-source comparison of COVID-19 framing across social media (Reddit), mainstream news, and government communication within a single analytical framework. Second, rather than inferring frames solely from lexical topics or predefined categories, we operationalize syntax-aware representations to examine how shared topical concepts are grammatically realized across sources. By linking quantitative divergence measures to interpretable grammatical evidence, our approach enables mixed-method framing analysis that captures both scale and linguistic nuance, clarifying how different communicators construct meaning and responsibility during a prolonged public health crisis.

Data

The analysis draws on a balanced corpus representing three communication sources within the U.S. information environment during the COVID-19 pandemic: public discourse, mainstream news media, and state-level government communication. To capture sustained patterns of framing beyond short-term reactions, all three corpora span the period from March 2020 through December 2022, covering the initial outbreak, multiple pandemic waves, and later phases of public health response.

News articles were collected from six major U.S. outlets selected to balance national and regional coverage. National outlets (*The New York Times*, *The Washington Post*, and *The Wall Street Journal*) represent widely circulated mainstream journalism with agenda-setting influence, while regional outlets in southeastern Pennsylvania (*The Philadelphia Inquirer*, *The Philadelphia Tribune*, and *Morning Call*) capture localized reporting and community-specific framing. All articles published between March 2020 and December 2022 were manually screened to ensure direct relevance to COVID-19, yielding 518 validated news articles.

To enable meaningful cross-source comparison and avoid dominance by any single source, the Reddit and government corpora were subsampled to align in scale with the curated news dataset. Reddit data were drawn from *r/COVID19* and *r/Coronavirus*, two large, topic-focused subreddits that hosted sustained public discussion throughout the pandemic. Posts published between March 2020 and December 2022 were retained if they were in English, had not been deleted, and exceeded a minimum length threshold of 20 tokens. Each Reddit data point aggregates an original post with its comments, preserving conversational context and framing dynamics. This procedure resulted in 549 Reddit discourse units.

Government communication consists of 570 state-level Department of Health (DoH) bulletins collected from 27 U.S. states. Bulletins were restricted to English-language content to ensure linguistic comparability. Several DoH portals reused standardized templates, so we applied a deduplication filter to remove near-duplicate documents. States without archived COVID-19 content, with insufficient separation of COVID-related materials, or with predominantly non-English publications were excluded.[1]

After tokenization, the initial corpus contains 154,580 tokens. Given substantial differences in document length, verbosity, and repetition patterns across sources, we applied a downsampling procedure as part of data preprocessing. This procedure was diagnostically guided by comparisons across multiple candidate thresholds. The goal of this step was to remove low-information and highly repetitive documents while preserving topical coverage across sources.

In practice, documents were flagged for removal based on a combination of document length, token repetition, and disproportionate concentration on source-specific or boilerplate terms (e.g., recurring disclaimers in government bulletins). Subsampling was further constrained to preserve the overall topic distribution within each source, ensuring that no topic was selectively removed from a particular source.

We evaluated multiple downsampling thresholds (0.1, 0.15, 0.2, and 0.25) and found that a threshold of 0.2 achieved the best balance between reducing source dominance and retaining informative content. Lower thresholds removed insufficient redundancy, while higher thresholds resulted in unnecessary information loss. After processing, the final corpus contains 154,178 tokens and substantially reduces extreme imbalances in document volume across sources.

As shown in Figure 1, downsampling leads to more comparable source contributions within each topic. This helps reduce the risk that downstream syntactic framing differences are driven by artifacts of source volume rather than

[1] Other states' DoH either did not archive their COVID-19 communications, did not clearly separate COVID-19 materials from other announcements, or primarily published content in languages other than English.

genuine communicative patterns.

Methodology

This section describes how we operationalize the syntactic outputs of the Linked Latent Theta Role (LLTR) model to analyze framing differences across communication sources. Our goal is not to modify or extend LLTR itself, but to construct interpretable and comparable representations of grammatical framing for government, mainstream news, and Reddit discourse.

Throughout the paper, we use the term *source* to refer to a communication subcorpus (i.e., government communications, mainstream news articles, and Reddit posts), rather than to individual outlets or authors.

LLTR: Syntactic Topic Modeling

We use the Linked Latent Theta Role (LLTR) model (Aghajari 2025), a syntactic topic model that jointly learns latent topics and the grammatical contexts in which topic words occur. LLTR associates each token occurrence with the dependency relations it participates in, yielding dependency-based evidence of the form (w, r, a), where w is a topic word, r is a dependency relation (e.g., *nsubj*, *amod*), and a is the linked argument. For example, the sentence "The virus spread rapidly" yields tuples such as (spread, nsubj, virus) and (spread, advmod, rapidly), which make explicit whether a topic word functions as an agent, patient, or modifier.

Why LLTR (vs. LDA and LDA-GR)

Our study adopts LLTR as a previously validated modeling interface for exploratory framing analysis, rather than proposing a new topic model or re-evaluating model performance. Prior work (Aghajari 2025) compared LLTR with LDA and LDA-GR through human evaluation by framing researchers and found that LLTR provides more coherent, connected, and interpretable contextual evidence for framing analysis. These advantages stem from LLTR's integration of topical co-occurrence and grammatical structure, whereas LDA-GR's strict word–relation binding increases sparsity and obscures less frequent but meaningful grammatical patterns. Accordingly, we use LLTR as the backbone for extracting syntactic evidence and focus our contribution on operationalizing its outputs for cross-source framing comparison.

Consistent with prior framing research, our analysis emphasizes interpretive grounding rather than statistical generalization. Framing interpretations in this work are supported by representative examples drawn directly from the underlying texts, which anchor each claim in observable linguistic patterns and ensure transparency and traceability of qualitative analysis.

Defining the Core Unit of Syntactic Framing

Topic words may appear in multiple grammatical roles across contexts and sources. For example, the same word can function as a subject in one construction, a modifier in another, or an object in a third. Consequently, knowing only that a topic word appears in a particular dependency relation (e.g., *nsubj* or *amod*) is insufficient for characterizing framing differences.

For framing analysis, we therefore treat the combination of a dependency relation and its linked argument, denoted as $(\mathit{reln}, \mathit{dep})$, as the core unit of syntactic evidence. This abstraction captures not only how a topic word is grammatically positioned, but also the semantic role it plays in relation to another lexical item.

Considering the dependency relation alone is too coarse for framing analysis. For example, while *nsubj* indicates that a word functions as a subject, distinguishing *virus–nsubj→spread* from *virus–nsubj→mutate* captures whether the narrative emphasizes contagion dynamics or biological evolution. Likewise, adjectival modification (*amod*) is only informative when interpreted relative to the noun it modifies: *political–amod→pressure* implies institutional constraint, whereas *political–amod→issue* frames the topic as contested or polarized, despite sharing the same relation label.

We do not explicitly model the governing head as a separate dimension. Instead, its framing contribution is implicitly captured through the dependency relation itself, which encodes how the topic word and the linked argument are structurally related. This design balances expressive power and interpretability, allowing us to compare grammatical framing across sources without introducing sparsity from fully lexicalized dependency triples.

Topic Count and Coherence

The number of topics was selected using the C_V topic coherence metric, which has been shown to correlate well with human judgments of topic interpretability (Röder, Both, and Hinneburg 2015). The C_V metric captures the degree to which the most probable words within a topic tend to co-occur in the corpus, with higher values indicating more semantically coherent and interpretable topics. We evaluated coherence scores for topic counts up to 20 and observed a local peak around eight topics. Lower topic counts tended to merge semantically distinct themes, while higher counts produced increasingly sparse and unstable grammatical patterns. Based on this diagnostic signal, we adopt eight topics for the analyses reported in this paper.

Source-Aware Enrichment of Grammatical Evidence

To highlight framing cues that are characteristic of each communication source, we apply a source-aware enrichment procedure to $(\mathit{reln}, \mathit{dep})$ pairs. For a given topic t, word w, source d, relation r, and dependent argument a, we compare the frequency of (r, a) within topic t and source d to its background frequency within the same source.

This filtering suppresses ubiquitous arguments that appear frequently across many topics and contexts regardless of framing. For example, arguments such as *people*, *cases*, or *health* occur with high frequency across government reports, news articles, and Reddit posts, but often convey little framing specificity on their own. Unless such arguments

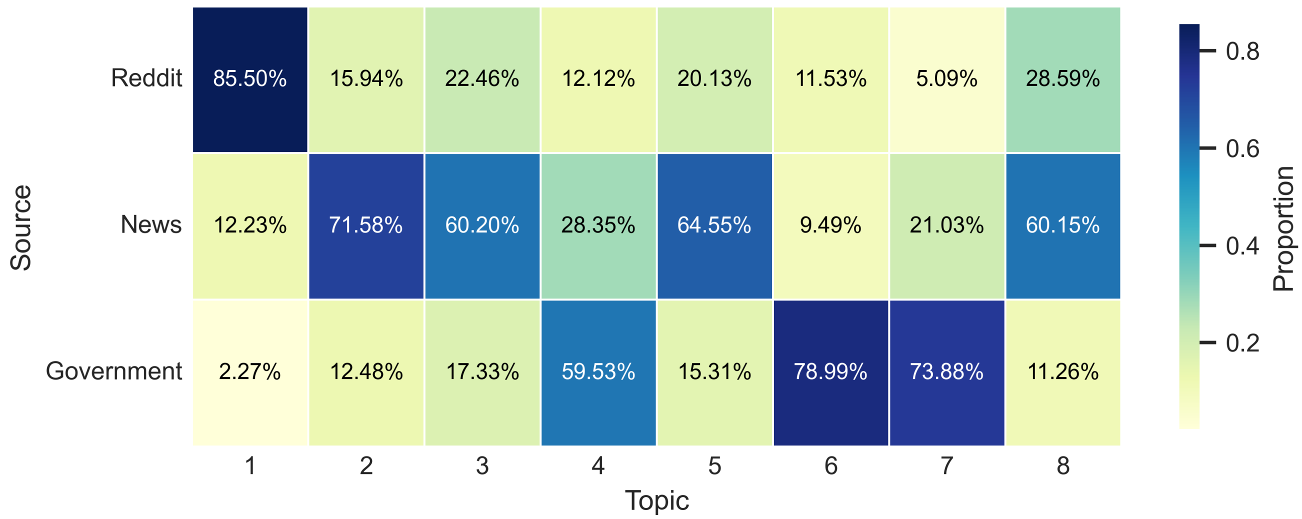

Figure 1: Proportion of documents from each source within each topic after downsampling. Downsampling reduces extreme source imbalances, producing more comparable source contributions across topics for cross-platform framing comparison; the full-resolution matrix appears in the Appendix.

are disproportionately associated with a particular topic and source, they are de-emphasized by the enrichment procedure.

Arguments with an enrichment score greater than 2 are retained and renormalized within each topic–word–source group. This step amplifies source-specific grammatical signals that are more likely to reflect framing differences rather than general topical salience.

Lexical Representation (Baseline)

To establish a lexical baseline for cross-source comparison, we also construct topic-word distributions for each topic and source. For a given topic, the lexical representation is obtained by reweighting the topic-level word distribution with source-specific observed usage of topic words. When a source exhibits no observed usage for a topic's vocabulary, the distribution defaults to the topic-level prior to ensure comparability across sources. These lexical distributions are used solely as a baseline and are not intended to capture framing differences.

Framing Vector Construction

For each topic, we select the top 10 most probable topic words for interpretability. A document contributes to a topic if its posterior probability for that topic is at least 0.35. Within each source, dependency counts are aggregated at the document level and normalized by document frequency to prevent long documents or highly active users from dominating the distributions.

Although topic-word selection limits lexical scope, framing differences are expressed through a larger set of grammatical realizations. We therefore construct framing vectors over the top $K = 100$ enriched $(reln, dep)$ pairs per topic, ranked by total enriched weight across sources. The value of K controls the granularity of grammatical evidence considered and is examined through robustness analyses in the Results section. This additional dimensionality preserves variation in how topic words are grammatically embedded, while ensuring comparability across sources. Reducing the representation to topic words alone would obscure framing differences that arise from grammatical structure.

Cross-Source Framing Divergence

Cross-source framing differences are quantified using Jensen–Shannon divergence (JSD) between source-specific framing distributions. We compute JSD under two representations: a lexical baseline based on topic–word distributions, and a syntactic representation based on enriched $(reln, dep)$ framing vectors. Framing vectors are normalized to probability distributions prior to comparison. For each topic, we report average pairwise JSD across the three sources to summarize overall framing separation. JSD is symmetric, bounded, and well-suited for comparing sparse probability distributions, making it appropriate for this setting.

Example Extraction for Interpretability

To link distributional patterns to natural language usage, we extract representative sentences illustrating high-weight $(reln, dep)$ pairs. Examples are drawn from documents with strong topic membership. Parser-confirmed dependencies are preferred when available; otherwise, co-occurrence matches are used. These examples provide qualitative grounding for the grammatical evidence analyzed in the Results section.

Results

We examine cross-source framing differences by comparing source-specific distributions of enriched grammatical evidence derived from dependency parses, as described above. For each topic, we examine framing by investigating the probability distribution over (*reln*, *dep*) pairs associated with topic words, aggregated within each source and normalized after source-aware enrichment. We quantify cross-source differences using Jensen–Shannon divergence (JSD), where higher values indicate greater separation in how sources grammatically frame shared topics.

We present results in three parts. First, we establish a lexical baseline showing that sources often use overlapping topic vocabularies. Second, we show that syntactic/relational divergence is substantially higher and varies by topic, indicating cross-source differences in how shared topical content is grammatically constructed. Third, we interpret what those syntactic differences mean for framing by inspecting representative high-weight grammatical evidence and example excerpts.

Table 1 provides an overview of the eight topics and summarizes the most salient source-conditioned framing tendencies for each topic. We reference this table throughout the Results section to contextualize which topics are being compared and what the dominant qualitative patterns look like within each source.

Lexical Similarity as a Baseline Across Sources

Across many topics, sources appear lexically similar, relying on overlapping topical vocabularies and distributing attention across topic words in comparable ways. This similarity establishes a baseline, indicating that cross-source differences are not primarily driven by vocabulary choice.

Figure 2 reports lexical JSD for each topic across each pair of sources. Lexical divergence is computed by constructing a topic-word distribution over the full topic vocabulary for each source. Specifically, we reweight the topic-level word distribution with the source-specific observed usage of those words. When a source exhibits no observed usage for a topic's vocabulary, the distribution falls back to the topic-level prior, ensuring the lexical JSD is defined for all source pairs. Overall, lexical divergence is relatively low, suggesting significant topical similarity among sources. However, as shown next, shared topical vocabulary does not necessarily imply similar framing.

Identifying Cross-Source Framing Differences with Syntactic Evidence

In contrast to lexical similarity, syntactic/relational divergence is substantially higher and varies by topic, indicating that sources often frame shared topical content differently by embedding the same topic words in distinct grammatical roles and relations.

Figure 3 summarizes syntactic/relational JSD by topic across government, news, and Reddit. Across all topics and source pairs, the mean divergence equals 0.751, indicating substantial cross-source separation in grammatical framing despite shared topical vocabulary. Notably, these values are consistently higher than the corresponding lexical JSD scores, reinforcing that the observed differences arise from how topics are grammatically constructed rather than from surface-level word choice. Divergence is not uniform across topics, suggesting that framing differences are topic-conditioned rather than source-invariant. Consistent with the

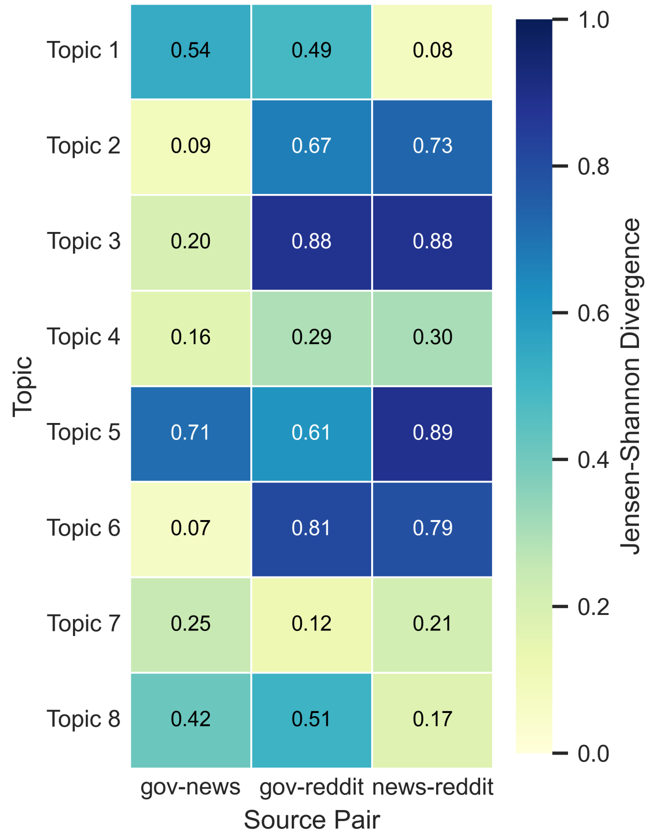

Figure 2: Lexical Jensen–Shannon divergence (JSD) by topic across each pair of sources. Lower values indicate that sources use similar topic vocabularies and distribute attention across topic words similarly. These results suggest **moderate to high topical similarity** among all three sources.

qualitative summaries in Table 1, we also observe that government and news are often closer to each other than either is to Reddit for several topics, though this alignment varies by topic.

Several topics exhibit strong pairwise alignment with a third source diverging. For example, Topic 4 (Vaccines & Vaccination) shows relatively low divergence between government and news (JSD = 0.24), while Reddit diverges more markedly from government (JSD = 0.77) and from news (JSD = 0.74). This pattern is consistent with shared institutional packaging of guidance in official communication and journalistic reporting, and more conversational, experience-centered realizations in Reddit discussion. In contrast, Topic 8 (Pandemic Timeline & Persistence) shows lower divergence between news and Reddit (JSD = 0.18) while government remains more distinct from either (JSD = 0.78 and JSD = 0.83, respectively), indicating convergence between journalistic and public framing for this topic.

Some topics exhibit high divergence across all source pairs. For instance, Topic 5 (Politics & Partisanship) exhibits consistently elevated divergence (JSD $\geq$ 0.97), suggesting

Topic & Keywords	Government	News	Reddit
T1 Public Discourse & Evaluation *really, good, coronavirus, people, virus, well, see*	Frames discussion through general public messaging and broad evaluative statements.	Embeds evaluative language within mediated reporting and attribution structures.	Centers conversational assessment, personal judgment, and informal evaluation.
T2 Economy & Disruption *market, economy, economic, business, jobs, companies*	Emphasizes losses, recovery, and financial response through administrative accounting.	Frames economic disruption via expert analysis and market interpretation.	Focuses on personal financial impact and everyday economic disruption.
T3 Community, Sports & Culture *teams, games, music, park, city, church*	Anchors activities in organized institutions and formal community roles.	Frames events through reporting of scheduled activities and public attendance.	Emphasizes local routines, coordination, and shared community experience.
T4 Vaccines & Vaccination *vaccine, vaccinated, doses, vaccination, disease, health*	Emphasizes eligibility, distribution, and procedural rollout of vaccination.	Frames vaccination through clinical attribution and expert mediation.	Centers personal experience, perceived risk, and individual decision-making.
T5 Politics & Partisanship *president, administration, biden, trump, congress, political*	Frames political action through institutional procedure and formal records.	Emphasizes conflict, criticism, and political accountability.	Highlights named actors, performative politics, and evaluative stance.
T6 Guidance, Testing & Cases *health, testing, symptoms, department, information, cases*	Stresses official guidance, reporting, and public health procedure.	Frames updates through mediated reporting and expert interpretation.	Focuses on access to information, uncertainty, and personal assessment.
T7 Aid, Programs & Public Support *support, aid, federal, state, program, governor*	Highlights assistance programs and eligibility as institutional provision.	Frames support through organizational response and consequence.	Emphasizes individual need, access, and first-hand requests.
T8 Pandemic Timeline & Persistence *still, months, time, first, last, pandemic*	Frames the pandemic through temporal progression and administrative phases.	Emphasizes continuity, change, and retrospective interpretation.	Focuses on lived duration, fatigue, and ongoing impact.

Table 1. Topic overview and source-specific framing summaries. Keywords are italicized under topic names.

that all three sources frame the topic in distinct ways. Note that this syntactic/relational divergence is independent of lexical divergence. To wit, Topic 6 (Health Guidance, Testing & Cases) has high lexical divergence between news and Reddit (JSD = 0.79), yet only moderate syntactic/relational divergence (JSD = 0.54). By contrast, Topic 7 (Aid, Programs & Public Support) shows comparatively lower divergence across sources (JSD$\leq$0.33), indicating more similar framing.

We compute syntactic JSD using the Top-100 (*reln*, *dep*) pairs per topic with uniform smoothing when a source exhibits no observed relational mass under the selected pair vocabulary. This avoids undefined comparisons and treats structural silence as maximum uncertainty under the restricted representation. As robustness checks, we also evaluated Top-50 and full-pair settings. Top-50 yields lower absolute divergence (because the representation is more compressed), while full-pair settings push divergence upward due to increased sparsity in very high-dimensional pair vocabularies. However, the relative topic-level patterns and the qualitative conclusions about which topics separate sources most strongly remain consistent.

Interpreting Framing Differences with Grammatical Evidence

To interpret what the distributional differences in Figure 3 mean for framing, we examine representative grammatical evidence associated with high-divergence topic words. We treat (*reln*, *dep*) configurations as *evidence* of how concepts are positioned in sentences rather than as frames themselves. This evidence helps explain how similar topical content can

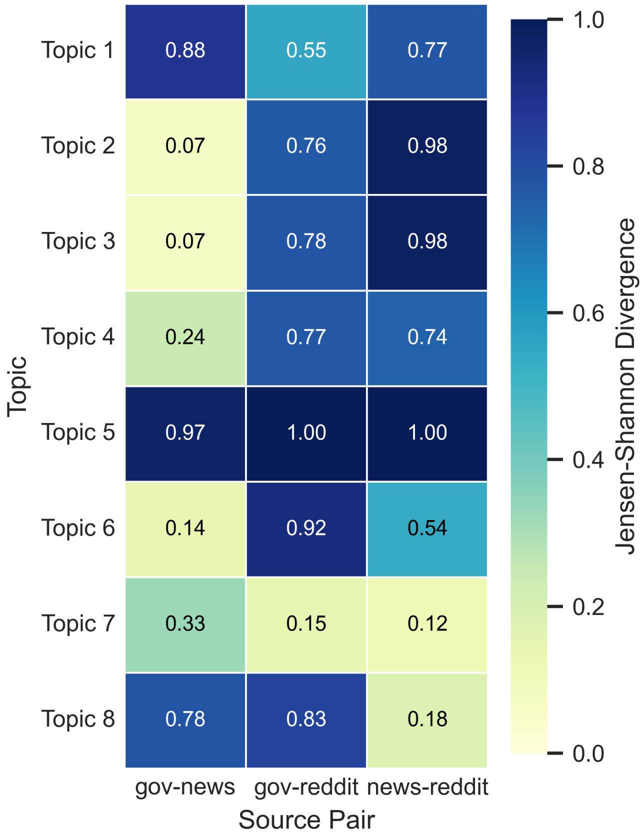

Figure 3: Syntactic/relational Jensen–Shannon divergence (JSD) by topic across each pair of sources (Top-100 pairs with uniform smoothing). Lower values indicate greater similarity in grammatical evidence of framing. These results suggest that **government and news are often more similar to each other than either is to Reddit**, though patterns vary by topic.

be packaged differently across sources, especially in terms of attributing agency and responsibility.

Across topics, a recurring pattern concerns how agency and responsibility are grammatically packaged across the three sources. Government texts tend to embed topic words in procedural and institutional structures, frequently surrounding them with temporal, prepositional, or administrative scaffolding that foregrounds eligibility, compliance, or formal action. News discourse more often embeds the same topic words in attribution and reporting structures, emphasizing expert mediation and evaluative narration. Reddit discourse more frequently realizes topic words in experiential or projected-action clauses and often includes evaluative modifiers, reflecting first-hand experience and stance.

For example, Topic 6 (Guidance, Testing & Cases) exhibits asymmetric levels of syntactic divergence across source pairs. Government and news discourse show very low divergence (JSD = 0.14), indicating close alignment in how guidance-related concepts are grammatically realized, whereas government and Reddit discourse diverge substantially (JSD = 0.92), with news and Reddit displaying a moderate level of divergence (JSD = 0.54).

In both government and news sources, guidance-related terms are predominantly realized through institutional directives and recommended actions. Government bulletins frequently contain constructions such as "Residents should follow department guidance ..." or "The department recommends measures to protect the public ...," where guidance is embedded in procedural clauses that foreground institutional authority and compliance. News reporting adopts closely aligned grammatical structures, often mediating these directives through attribution while preserving their institutional orientation.

In contrast, Reddit discussion more often frames the same guidance-related concepts through experiential and interpretive statements, such as "I'm not sure what the guidance means for my situation ...," shifting emphasis toward individual judgment and decision-making. This example shows how Reddit discourse emphasizes individual agency and experience—how to decide what I do for "my situation"—while government discourse emphasizes compliance with institutional authority, with what "the department recommends."

This pattern indicates that Topic 6 corresponds to an intermediate level of divergence: two sources exhibit close alignment, while the third diverges in systematic and interpretable ways.

By contrast, Topic 5 (Politics & Partisanship) represents an extreme case, exhibiting consistently high syntactic divergence across all source pairs (JSD $\geq$ 0.97). This pattern indicates that government, news, and Reddit discourse each frame political content in structurally distinct ways. One illustrative keyword within this topic is *political*, which appears across all three sources but is embedded in markedly different grammatical configurations with divergent framing implications.

In government discourse, *political* most frequently modifies nouns such as *pressure*, often in constructions that oppose political influence to institutional rationality or scientific authority (e.g., "We must follow the science, not political pressure ..."). Grammatically, this pairing frames politics as an external force that should be excluded from legitimate decision-making, reinforcing a normative boundary between evidence-based governance and partisan interference.

News discourse, in contrast, tends to embed *political* in evaluative noun phrases such as "political issue" or "divisive political issue." These constructions frame politics as a site of controversy or polarization, emphasizing conflict and disagreement without clearly assigning responsibility or legitimacy. Reddit discourse diverges further still, frequently associating *political* with nouns such as *theater* or *games*, producing a framing characterized by cynicism and performative critique informed by perceived irresponsibility among political actors. Although the lexical item *political* remains constant, its grammatical role and linked arguments vary substantially across sources, accounting for the uniformly high syntactic divergence observed for this topic.

Conversely, topics with lower divergence exhibit similar distributions of grammatical evidence, reflecting alignment in how sources package actions, actors, and consequences. For example, Topic 7 (Aid, Programs & Public Support) exhibits the lowest divergence across the three sources (JSD ≤ 0.33). Within this topic, one salient topic word is *state*, which is realized in highly similar grammatical roles across government, news, and Reddit discourse. Government communication typically embeds *state* in administrative and programmatic constructions (e.g., "the state provides assistance through ..."), news articles adopt comparable institutional framing when reporting on funding or policy actions (e.g., "the state announced additional support ..."), and Reddit discussion likewise refers to the state as a background institutional entity in discussions of aid and eligibility (e.g., "the state is offering help for ..."). Across sources, *state* is grammatically packaged as an institutional locus of action rather than as a site of interpretation, contestation, or evaluation, producing closely aligned role–argument distributions and correspondingly low syntactic divergence.

Such patterns, visible at the level of individual keywords, help clarify what drives patterns of high and low JSD. These differences illustrate how syntactic divergences—between subject and object realization, modification and complementation, etc.—serve as evidence for differences in framing, especially institutional versus experiential accounts.

Overall, these findings demonstrate that cross-source framing differences arise from grammatical choices that condition how meaning is constructed, rather than from surface-level topic overlap. By quantifying divergence over grammatical evidence and interpreting the structures that drive it, the analysis shows how public, media, and government discourse systematically frame shared issues in different ways.

Discussion

This study shows that cross-source framing differences are not reducible to differences in topical emphasis, but emerge from systematic variation in how shared concepts are grammatically realized across communication sources. By analyzing distributions of syntactic evidence associated with common topic keywords, we show that public, media, and governmental discourse differ in framing, especially in terms of how they assign agency and responsibility. Our contribution lies not in proposing a new model, but in operationalizing grammar-aware representations for comparative framing analysis at scale and in the substantive framing insights enabled by this operationalization. This section places that contribution in conversation with related prior work to draw out potential implications.

Grammatical Structure as a Mechanism of Framing

A key implication is that framing functions are enacted through recurrent grammatical choices that shape how events, actors, and responsibilities are presented. While framing theory emphasizes selection and salience (Entman 1993; Scheufele 1999; Chong and Druckman 2007), our results illustrate how these functions materialize in concrete syntactic patterns. Across topics, government communication tends to package events through institutional and procedural constructions (foregrounding administrative action, eligibility, and compliance), news discourse more often embeds the same concepts in attributional or evaluative structures (emphasizing expert judgment and accountability), and Reddit discourse frequently realizes them in experiential or projected-action clauses (linking events to personal consequence and affect). These source-conditioned realizations help explain why shared topics can produce distinct narrative effects even when lexical overlap is substantial.

Implications for Computational Framing Research

Our findings underscore the limits of approaches that rely solely on lexical features or topic distributions to study framing. This point runs contrary to prior work suggesting that lexical features alone are sufficient for identifying and analyzing framing (Baumer et al. 2015; Naderi and Hirst 2017; Walter and Ophir 2019). Shared words such as *virus*, *cases*, or *vaccines* appear across government, news, and social media, yet their grammatical realizations reveal whether they are framed as active threats, administrative counts, or objects of personal concern—distinctions that topic-only or bag-of-words representations often obscure (Nicholls and Culpepper 2021; Ali and Hassan 2022). By treating grammatical role–argument configurations as interpretable evidence rather than frames themselves, our approach complements prior computational work on agenda setting, attention, and sentiment in crisis communication (Cinelli et al. 2020; Han, Yang, and Piterou 2021), while providing finer-grained access to how narratives assign agency and responsibility within shared topical contexts without requiring supervised frame labels.

Comparison with Other COVID-19 Framing Analyses

In comparison with prior analyses of COVID-19 framing, most of which focus on a single type of communication source, our results both reinforce and complicate existing accounts of how pandemic-related narratives are constructed across platforms.

Consistent with prior studies of social media discourse, we find that Reddit discussions prominently exhibit experiential and vernacular framing, characterized by emotionally charged and first-person narratives (Wicke and Bolognesi 2020; Tsao et al. 2021). Likewise, our observations of institutional and procedural framing in government communication align with earlier analyses emphasizing guidance, collective responsibility, and compliance-oriented messaging in public health sources (Kandzer et al. 2022; Rao et al. 2020; Song et al. 2025). These convergences with prior work support the validity of our syntactic, grammar-aware approach to identifying framing patterns.

However, our findings diverge from several prior studies of news media framing, which report a strong emphasis on fear, human interest, and experiential narratives in pandemic coverage (Ogbodo et al. 2020; Ebrahim 2022; Rooke 2021). In contrast, our analysis suggests that, over the full span of the pandemic, mainstream news discourse more closely resembles the institutional framing patterns observed in government communication, while strongly experiential framings are largely concentrated in Reddit discussions. One plausible explanation for this difference is temporal scope: whereas many prior analyses focus on early-pandemic coverage, our dataset spans more than two and a half years of COVID-19 reporting, capturing later phases in which news framing may have shifted toward routinized, policy-oriented, and administratively mediated narratives.

Beyond these substantive comparisons, our findings also extend prior grammar-aware and role-based computational approaches. Previous work has incorporated syntactic information into topic models to improve coherence or capture selectional preferences (Boyd-Graber and Blei 2008; Delpisheh and An 2014; Ritter, Mausam, and Etzioni 2010), and more recent studies have leveraged predicate–argument structure to examine narrative dynamics in online discourse (Zhao et al. 2024). Our results show that grammatical structure is not only useful for modeling narrative or topical organization, but also provides a systematic basis for comparing framing across heterogeneous communication sources. In particular, we find that differences in grammatical role–argument configurations account for substantial cross-source framing divergence even when topical vocabularies overlap, highlighting local syntactic choice as a key mechanism through which framing differences are enacted.

Limitations and Scope

This study relies on qualitative interpretation of high-divergence grammatical evidence and representative examples to explain framing differences. Alternative readings of the same syntactic patterns are possible, and future work could incorporate human validation to assess interpretive robustness. In addition, balancing state-level health communication with other sources required downsampling, which may reduce coverage of less frequent themes. While we do offer some insights about variations in COVID-19 discourse, the primary contribution involves demonstrating how syntactic framing analysis can reveal systematic cross-source differences. Our approach also inherits limitations from dependency parsing; future work could incorporate parser-confidence estimation, phrase mining, or targeted analyses of specific constructions (e.g., agentless passives or modal verbs), alongside emerging computational framing approaches (Nicholls and Culpepper 2021; Ali and Hassan 2025).

Future Directions

First, grammar-aware framing analysis could be extended to richer semantic representations, such as semantic role labeling (SRL). Such an approach would transcend syntactic relations to explicitly model predicate–argument structure and abstract participant roles (e.g., agent, patient, beneficiary) (Palmer, Gildea, and Xue 2010). Integrating SRL with topic modeling may enable analyses of how higher-level semantic roles interact with topical framing across sources, complementing the dependency-based evidence used here and building on work that leverages predicate–argument structure to study collective narratives in online discourse (Zhao et al. 2024). Second, human-in-the-loop evaluation could assess how analysts interpret grammatical framing cues and validate the robustness of syntactic divergence measures. Finally, applying this framework to other crises, political events, or multilingual settings could clarify how institutional, media, and public framing strategies generalize across contexts and languages.

Conclusion

This study demonstrates that cross-source framing differences cannot be explained by topical variation alone, but are reflected in how shared topical concepts are grammatically realized across communication sources. Although government, news, and Reddit discourse often rely on overlapping topic vocabularies, our results show that these shared lexical resources are embedded in systematically different grammatical configurations.

By comparing lexical and syntactic representations directly, we find that lexical similarity frequently co-exists with substantial divergence in syntactic-based evidence. This divergence varies by topic and source pair, indicating that framing differences are conditioned by how topic words participate in grammatical relations and argument structures, rather than by the presence or absence of particular keywords. Linking these quantitative divergence patterns to representative examples illustrates how similar topical content can be packaged differently across institutional, journalistic, and public discourse.

The contribution of this work lies not in introducing a new topic model, but in demonstrating how grammar-aware topic representations can be operationalized for comparative framing analysis in a way that is both scalable and interpretable. By treating grammatical role–argument configurations as framing evidence rather than as frames themselves, the proposed framework enables systematic comparison of how meaning is constructed across heterogeneous communication environments.

While grounded in COVID-19 discourse, this approach generalizes to other settings where institutional, media, and public narratives intersect. It offers a computationally tractable pathway for studying framing beyond surface lexical similarity, and for connecting large-scale distributional patterns to qualitative interpretation of how shared issues are linguistically constructed across sources.

Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under Grant Number IIS-2212265. Thanks to Lillian Mauger, Chase Mattingly, and Isabel Koval for assistance with data collection. We also thank Chenchen Mao, Laila Jama, and the anonymous reviewers for their constructive feedback on earlier versions of this paper.

References

Aghajari, Z. 2025. *Adopting an Ecological Approach to Misinformation: Understanding the Broader Scope and Impacts of Misinformation on Online Communities*. Ph.D. thesis, Lehigh University.

Ali, M.; and Hassan, N. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, 9335–9348.

Ali, M.; and Hassan, N. 2025. Semantic-based Unsupervised Framing Analysis (SUFA): A Novel Approach for Computational Framing Analysis. *arXiv preprint arXiv:2505.15563*.

Baumer, E.; Elovic, E.; Qin, Y.; Polletta, F.; and Gay, G. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics: human language technologies*, 1472–1482.

Bergenfeld, I. 2024. What can public health communicators learn from Reddit? A perspective for the next pandemic. *Frontiers in Public Health*, 12: 1348095.

Boyd-Graber, J. L.; and Blei, D. M. 2008. Syntactic Topic Models. In *Advances in Neural Information Processing Systems 21*, 185–192.

Chen, E.; Lerman, K.; and Ferrara, E. 2020. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance*, 6(2): e19273.

Chong, D.; and Druckman, J. N. 2007. Framing theory. *Annual Review of Political Science*, 10: 103–126.

Cinelli, M.; Quattrociocchi, W.; Galeazzi, A.; Valensise, C. M.; Brugnoli, E.; Schmidt, A. L.; Zola, P.; Zollo, F.; and Scala, A. 2020. The COVID-19 social media infodemic. *Scientific reports*, 10(1): 16598.

Delpisheh, E.; and An, A. 2014. Topic modeling using collapsed typed dependency relations. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 146–161. Springer.

Ebrahim, S. 2022. The Corona Chronicles: Framing Analysis of Online News Headlines of the COVID-19 Pandemic in Italy, USA and South Africa. *Health SA Gesondheid*, 27(0): a1683.

Entman, R. M. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4): 51–58.

Hallahan, K. 1999. Seven models of framing: Implications for public relations. *Journal of public relations research*, 11(3): 205–242.

Han, C.; Yang, M.; and Piterou, A. 2021. Do news media and citizens have the same agenda on COVID-19? An empirical comparison of Twitter posts. *Technological Forecasting and Social Change*, 169: 120849.

Hayek, L. 2024. Media Framing of Government Crisis Communication During Covid-19. *Media and Communication*, 12: 7774.

Holmes, I. D. 2025. Framing COVID-19: How UK Government and Media Narrated the "Crisis". *Politics & Policy*, 53(3): e70040.

Kandzer, M.; Castano, V.; Baker, L. M.; and McLeod-Morin, A. 2022. Framing friction: a content analysis investigating how the CDC framed social media communication with the public during the COVID-19 pandemic. *Journal of Applied Communications*, 106(1): 4.

Kermani, H.; Makou, A. B.; Tafreshi, A.; Ghodsi, A. M.; Atashzar, A.; and Nojoumi, A. 2024. Computational vs. qualitative: analyzing different approaches in identifying networked frames during the Covid-19 crisis. *International Journal of Social Research Methodology*, 27(4): 401–415.

Mohammadi, E.; Tahamtan, I.; Mansourian, Y.; and Overton, H. 2022. Identifying frames of the COVID-19 infodemic: thematic analysis of misinformation stories across media. *JMIR infodemiology*, 2(1): e33827.

Naderi, N.; and Hirst, G. 2017. Classifying Frames at the Sentence Level in News Articles. In *RANLP 2017 – Recent Advances in Natural Language Processing Meet Deep Learning*, 536–542. Incoma Ltd.

Nicholls, T.; and Culpepper, P. D. 2021. Computational identification of media frames: strengths, weaknesses, and opportunities. *Political Communication*, 38(1-2): 159–181.

Nienhaus, S.-M. 2024. Framing the COVID-19 pandemic: A comparison of political and media framing during the first three waves in Germany. *German Politics*, 33(4): 740–763.

Ogbodo, J. N.; Onwe, E. C.; Chukwu, J.; Nwasum, C. J.; Nwakpu, E. S.; Nwankwo, S. U.; Nwamini, S.; Elem, S.; and Ogbaeja, N. I. 2020. Communicating health crisis: a content analysis of global media framing of COVID-19. *Health promotion perspectives*, 10(3): 257.

Palmer, M.; Gildea, D.; and Xue, N. 2010. *Semantic Role Labeling*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Panzeri, F.; Di Paola, S.; and Domaneschi, F. 2021. Does the COVID-19 war metaphor influence reasoning? *Plos one*, 16(4): e0250651.

Rao, H. R.; Vemprala, N.; Akello, P.; and Valecha, R. 2020. Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. *International Journal of Information Management*, 55: 102187.

Reveilhac, M. 2022. The deployment of social media by political authorities and health experts to enhance public information during the COVID-19 pandemic. *SSM - Population Health*, 19: 101165.

Ritter, A.; Mausam; and Etzioni, O. 2010. A Latent Dirichlet Allocation Method for Selectional Preferences. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 424–434.

Röder, M.; Both, A.; and Hinneburg, A. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408.

Rooke, M. 2021. Alternative media framing of COVID-19 risks. *Current Sociology*, 69(4): 584–602.

Scheufele, D. A. 1999. Framing as a theory of media effects. *Journal of Communication*, 49(1): 103–122.

Song, T.; Yu, P.; Yecies, B.; Ke, J.; and Yu, H. 2025. Social media crisis communication and public engagement during COVID-19 analyzing public health and news media organizations' tweeting strategies. *Scientific Reports*, 15(1): 18082.

Sussman, K. L.; Bouchacourt, L.; Bright, L. F.; Wilcox, G. B.; Mackert, M.; Norwood, A. S.; and Allport Altillo, B. S. 2023. COVID-19 topics and emotional frames in vaccine hesitation: A social media text and sentiment analysis. *Digital Health*, 9: 20552076231158308.

Tahamtan, I.; Potnis, D.; Mohammadi, E.; Miller, L. E.; and Singh, V. 2021. Framing of and attention to COVID-19 on Twitter: thematic analysis of hashtags. *Journal of Medical Internet Research*, 23(9): e30800.

Tsao, S.-F.; Chen, H.; Tisseverasinghe, T.; Yang, Y.; Li, L.; and Butt, Z. A. 2021. What social media told us in the time of COVID-19: a scoping review. *The Lancet Digital Health*, 3(3): e175–e194.

Walter, D.; and Ophir, Y. 2019. News Frame Analysis: An Inductive Mixed-method Computational Approach. *Communication Methods and Measures*, 13(4): 248–266.

Wicke, P.; and Bolognesi, M. M. 2020. Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PloS one*, 15(9): e0240010.

Zhao, W.; Guo, S.; Lerman, K.; and Ahn, Y.-Y. 2024. Discovering collective narratives shifts in online discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 1804–1817.

Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, our research analyzes publicly available textual data from Reddit, news articles, and government bulletins. All content is anonymized and aggregated, with no personally identifiable information included.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes. We explicitly state that LLTR is not a contribution of this paper. Our contributions are (i) a systematic method that operationalizes an existing syntactic topic model for framing analysis, and (ii) empirical findings that show how grammatical structures differ across government, news, and Reddit discourse.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, the Methods section details the LLTR model, source-aware enrichment, and dependency-based framing vector pipeline that support our framing divergence claims.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see the Downsampling and Limitations sections. We address potential biases from boilerplate government language and informal Reddit grammar.
   (e) Did you describe the limitations of your work? Yes, the Discussion section covers limitations including parsing errors, English-only data, downsampling tradeoffs, and observational design.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, in the Discussion and Ethics sections, we note that framing diagnostics could be misused for manipulative messaging, and propose mitigation strategies.
   (g) Did you discuss any potential misuse of your work? Yes, we explicitly state that the LLTR outputs are not designed for optimizing persuasion and advocate for ethical oversight in downstream use.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we release only aggregated outputs and enriched vectors. No raw user-level data is shared. All modeling and filtering steps are described for reproducibility.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, our data use is restricted to public sources, and all content has been filtered for PII and offensive material.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, the methodology assumes topic posterior thresholds (≥ 0.35), dependency parse validity, and source-specific enrichment criteria to reflect framing.
   (b) Have you provided justifications for all theoretical results? Yes, our quantitative framing divergence results (JSD) are empirically derived and justified through vector construction and enrichment.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, we compare syntax-aware framing models to prior keyword-based approaches and discuss alternative explanations for framing divergence.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, we explicitly discuss how lexical overlap cannot explain the observed divergence, which instead arises from syntactic role and argument realization.
   (e) Did you address potential biases or limitations in your theoretical framework? Yes, see the Limitations section, where we address parsing errors, threshold effects, and source-specific document structure biases.

(f) Have you related your theoretical results to the existing literature in social science? Yes, we build on framing theory (Entman 1993; Hallahan 1999) and connect our findings to established work in crisis communication and political framing.

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science source? Yes, in the Discussion we highlight implications for public health communication, media literacy, and platform-specific messaging strategies.

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA. This study does not include formal theoretical proofs.

(b) Did you include complete proofs of all theoretical results? NA.

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, we include topic-word lists, enriched frame vectors, and dependency-based sentence examples in the Appendix and supplemental materials.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, we describe the LLTR configuration, topic numbers, coherence scores, thresholds for posterior inclusion, and enrichment filtering criteria.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? No, because our model uses deterministic parsing and topic assignments without random initialization.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, because the experiments ran on CPU without significant computational resources.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, we use Jensen–Shannon divergence to evaluate cross-source divergence in grammatical framing.

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA. The model is not used for classification or decision-making under uncertainty.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, without compromising anonymity...

(a) If your work uses existing assets, did you cite the creators? Yes, Reddit and DoH sources are referenced, and news data is contextualized in the corpus description.

(b) Did you mention the license of the assets? No, but all data are publicly available; Reddit data is used under platform terms, and government documents are public sources.

(c) Did you include any new assets in the supplemental material or as a URL? Yes, we include enriched topic–word–frame distributions and case study tables in the supplemental material.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA. All data are public and do not involve direct interaction with human subjects.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, we filtered for sensitive content and ensured no user names or PII appear in outputs.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? No, as our focus is on analysis, not dataset curation, though we provide enriched data summaries in support of reproducibility.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? No, though we describe each corpus's source and processing method in the Methodology.

6. Additionally, if you used crowdsourcing or conducted research with human subjects, without compromising anonymity...

(a) Did you include the full text of instructions given to participants and screenshots? NA. No human participants were involved.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.

(d) Did you discuss how data is stored, shared, and deidentified? Yes, all released data are deidentified, aggregated, and restricted to syntactic representations without raw text.

Appendix

Additional Methods and Implementation Details

This appendix provides implementation details to support reproducibility. Topic words are selected by model probability within each topic. Documents enter topic-level estimation when their posterior topic probability is at least 0.35. For each topic–word–source triple, (*Reln*, *Dep*) counts are aggregated at the document level and normalized by document frequency to limit the influence of high-volume outliers.

Source-specific enrichment follows the procedure described in the methodology section. For each source, background frequencies are estimated over the full corpus. An enrichment ratio is computed for each (*Reln*, *Dep*) pair, and pairs with enrichment $E > 2$ are retained. Retained pairs are re-normalized within each topic–word–source group.

For cross-source comparison, framing vectors are aligned on the top-$K = 100$ enriched (*Reln*, *Dep*) pairs per topic. All vectors are normalized to probability distributions. Jensen–Shannon divergence (JSD) is used as the sole metric for cross-source comparison throughout the paper. Small additive smoothing is applied to avoid zero-probability issues. All reported divergence values are computed on these normalized distributions using identical thresholds and vocabularies across sources.

Sentence examples are drawn from documents with topic posterior at least 0.35. Dependency relations are parser-confirmed when available; otherwise, co-occurrence matches are retained as weak evidence and marked accordingly in the supplemental files.

Sensitivity and Robustness

We evaluate the stability of the reported findings with respect to key hyperparameters. The enrichment threshold was varied in $\{1.5, 2, 3\}$. The top-K vocabulary size was varied in $\{50, 100, 150\}$. The topic posterior threshold was varied in $\{0.30, 0.35, 0.40\}$.

Across all settings, the ranking of topics by average JSD and the relative alignment patterns between sources remained stable. The same high-divergence topics and keywords consistently exhibited distinct grammatical evidence across sources. These results indicate that the reported framing differences do not hinge on a particular parameter choice.

Source-Conditioned Grammatical Framing

This appendix reports the full set of keyword-level grammatical evidence that underlies the quantitative divergence patterns discussed in the main text. For each topic, one to two high-salience keywords are selected based on cross-source divergence. For each selected keyword and source, the two most weighted grammatical evidence pairs are listed together with one representative sentence excerpt.

The tables are intended to support transparency and verification rather than to introduce additional interpretation. Grammatical relations are reported as (*Reln*, *Gov*, *Dep*) triples and serve as *framing evidence*, not as frames themselves. Interpretation of these patterns is provided in the Results section of the main paper.

Topic	Keyword	Source	Top grammatical evidence	Example excerpt
1	coronavirus	gov	*nn*(pandemic, coronavirus); *prep_of*(spread, coronavirus)	"The coronavirus pandemic has created new challenges for public health officials . . . "
1	coronavirus	news	*nn*(pandemic, coronavirus); *nn*(outbreak, coronavirus)	"As the coronavirus pandemic continues to impact the global economy, supply chains remain strained . . . "
1	coronavirus	reddit	*nn*(pandemic, coronavirus); *nn*(crisis, coronavirus)	"I'm so tired of this coronavirus pandemic ruining everything for us . . . "
1	virus	news	*nsubj*(virus, spread); *amod*(virus, chinese)	"The virus spread rapidly across the country, mostly among unvaccinated groups . . . "
1	virus	reddit	*nsubj*(virus, spread); *amod*(virus, chinese)	"We knew the virus would spread if people didn't take precautions . . . "
2	market	gov	*prep_in*(housing, market); *nn*(labor, market)	"Conditions in the housing market have tightened significantly . . . "
2	market	news	*nn*(stock, market); *nn*(labor, market)	"The stock market rallied today on news of the vaccine . . . "
2	market	reddit	*nn*(stock, market); *nn*(free, market)	"It feels like the stock market is completely detached from reality . . . "
3	city	gov	*nn*(york, city); *nn*(kansas, city)	"Reports from New York City indicate a decline in cases . . . "
3	city	news	*nn*(york, city); *nn*(kansas, city)	"New York City officials are planning to reopen schools . . . "
3	city	reddit	*nn*(york, city); *nn*(inner, city)	"Living in New York City during this has been a nightmare . . . "
4	health	gov	*prep_of*(department, health); *nn*(department, health)	"The Department of Health has issued new guidance . . . "
4	health	news	*nn*(officials, health); *nn*(experts, health)	"Public health officials warn of a winter surge . . . "
4	health	reddit	*nn*(issues, health); *nn*(care, health)	"Mental health issues are being ignored in this pandemic . . . "
5	political	gov	*amod*(science, political); *amod*(pressure, political)	"We must follow the science, not political pressure . . . "
5	political	news	*amod*(issue, political); *amod*(gain, political)	"Masks have become a divisive political issue . . . "
5	political	reddit	*amod*(theater, political); *amod*(games, political)	"This is all just political theater by the governor . . . "
5	administration	gov	*nn*(Biden-Harris, administration); *prep_on*(commitment, administration)	"HHS worked to deliver on the Biden-Harris Administration's commitment to tackle the COVID-19 pandemic . . . "
5	administration	news	*nn*(Trump, administration); *nsubj*(administration, Trump)	"The Trump administration announced new measures earlier this week . . . "
5	administration	reddit	*prep*(administration, commission); *prep*(administration, state)	"People keep blaming the state administration for how the commission handled the guidelines . . . "
6	cases	gov	*nn*(cases, reported); *amod*(cases, new)	"The total number of reported cases stands at . . . "
6	cases	news	*nn*(coronavirus, cases); *amod*(cases, new)	"Coronavirus cases are spiking in the south . . . "
6	cases	reddit	*nn*(covid, cases); *nn*(cases, rising)	"Why are covid cases still rising despite the vaccine . . . "
7	state	gov	*prep_in*(department, state); *nn*(health, state)	"Contact your state health department for vaccine sites . . . "
7	state	news	*nn*(health, state); *nn*(officials, state)	"State officials have declared a state of emergency . . . "
7	state	reddit	*nn*(lines, state); *prep_in*(live, state)	"Traveling across state lines is confusing with these rules . . . "
8	people	gov	*nsubj*(vaccinated, people); *dobj*(get, people)	"We urge all people to get vaccinated immediately . . . "
8	people	news	*nsubj*(vaccinated, people); *amod*(people, young)	"Many young people are hesitant to take the shot . . . "
8	people	reddit	*nsubj*(think, people); *nsubj*(say, people)	"People think this is over but it's really not . . . "

Table 2. Complete source-conditioned grammatical evidence for high-salience keywords across all eight topics. Evidence is reported as *Reln(Head, Dep)* pairs and provides the linguistic grounding for the qualitative case studies discussed in the Results section.

Quantitative Complements (Full-Resolution Figures)

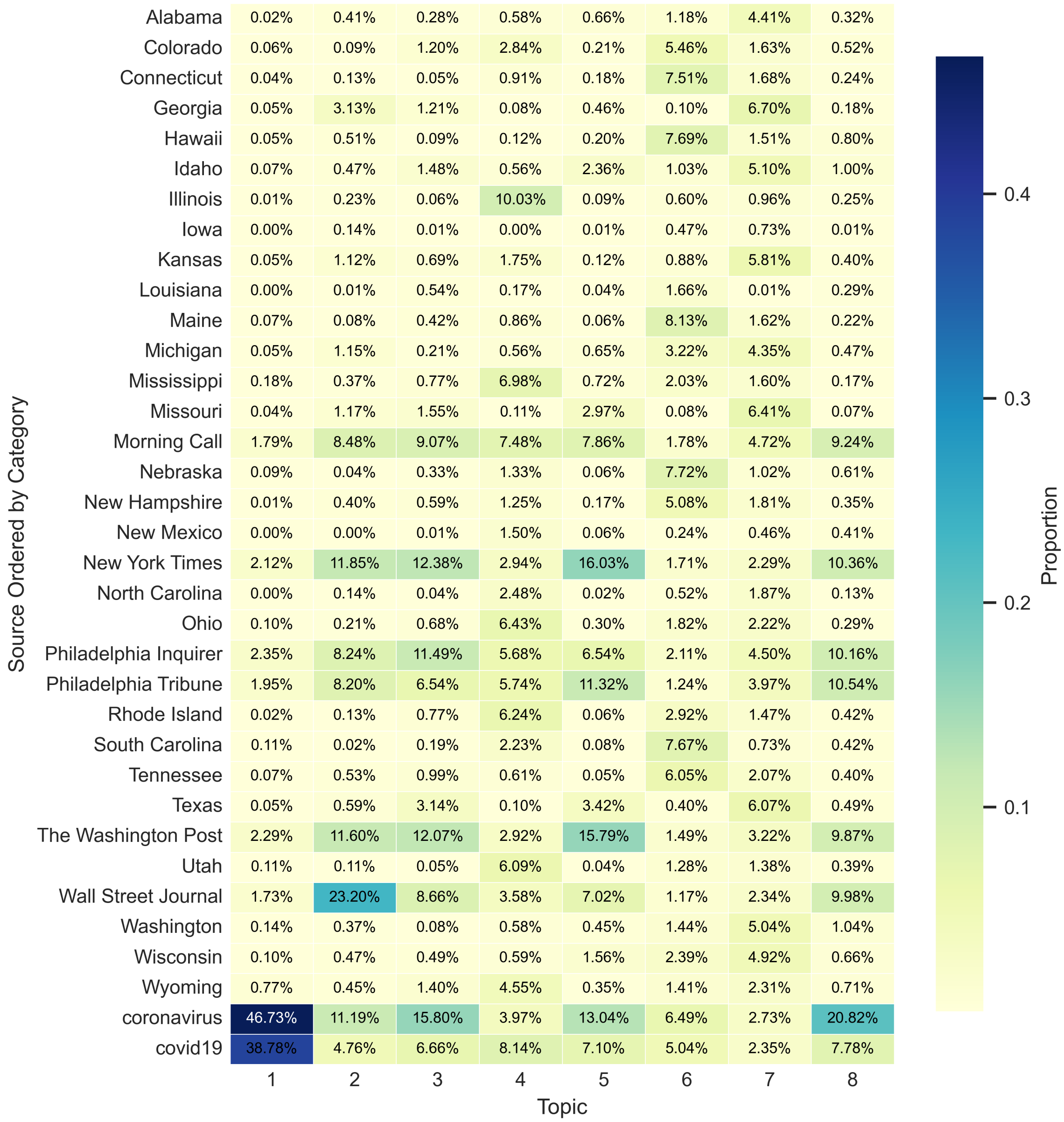

Figure 4: Proportion of documents from each source within each topic after downsampling (full resolution). Sources include all state-level Department of Health (DoH) bulletins, manually screened mainstream news articles, and two subreddits (*r/COVID19* and *r/Coronavirus*). This figure provides the complete source-by-topic matrix used to verify that topic estimates are not dominated by any single high-volume source.