CCF Database: A Machine-Learning-Annotated Corpus of 266,271 Canadian Climate Articles (1978–2024)

Antoine Lemor[1], Alizée Pillod[2], and Matthew Taylor[2]

[1]Université de Sherbrooke
[2]Université de Montréal

Abstract

Climate discourse does not merely mirror public debate around climate change, but actively shapes how societies perceive and respond to the crisis. Yet large-scale, fine-grained analysis of climate discourse remains hindered by fragmented corpora and ad hoc coding schemes. Here, we introduce the Canadian Climate Framing (CCF) database, a generalizable annotation framework, and training data to overcome these limitations. The CCF comprises 266,271 articles from 20 Canadian newspapers (1978–2024), processed into 9.2 million bilingual sentences (82.9% English, 17.1% French) with 65 hierarchical annotations: eight thematic frames, nine actor types, eight event categories, solution strategies, emotional tone, geographic focus, and named entities. Construction relied on transformer classifiers (BERT/CamemBERT) trained through human-in-the-loop iteration on 4,000+ expert-coded sentences, validated against a gold standard (F1=0.866) with confirmed intercoder reliability. Four analytical applications—linking political figures to science skepcticism, mapping discourse polarization, modeling framing editorial prioritization, and analysing epistemic authority networks—illustrate the database's research potential. Both the CCF and its underlying framework and training data are designed for replication across national and linguistic contexts.

1 Background & Summary

On October 18, 1983, *The Globe and Mail* ran a story headlined: "Disastrous warming of Earth to start in '90s, report warns." The article included the following warning:

> "In the next century, [the report] warns, the world will have to learn to cope with major changes in climate patterns, with disrupted food production and with significantly higher coastal waters."

This excerpt shows that climate science has been present in Canadian media coverage for more than forty years. But more fundamentally, it tends to confirm the premise that news coverage offers a window onto social and political realities. Because most citizens do not have direct access to academic research, they rely largely on media coverage to make sense of climate change. As a result, news media occupies a central position in the climate problem: they shape not only what people think about, but also how they think about it [1]. Beyond their impact on public opinion, media are also widely seen as influencing political agendas, even if the magnitude of this influence varies across issues and contexts [2]. Consequently, the media are not simply a neutral arena for public debate; they actively participate in defining and constructing the very problems under discussion [3, 4].

This paper introduces the *Canadian Climate Framing (CCF)* database, a machine-learning-preprocessed corpus of 266,271 climate-related newspaper articles published between 1978 and 2024 across 20 major Canadian media outlets. The dataset contains more than 9,198,158 sentences, all annotated within a 65-category hierarchical coding scheme. By focusing on Canadian print media, this study captures both a historically influential source of climate information[1] and a uniquely complex national context[2], providing, to the authors' knowledge, a rich case study for understanding how media coverage intersects with regional, economical, linguistic, and/or political differences.

Systematic analysis of climate communication in Canada and elsewhere has long been constrained by several methodological challenges: until recently, large-scale corpus collection and precise, reproducible content analysis were virtually impossible. Table A in Appendix provides an overview of prior studies in this area, highlighting the scope, focus, and coverage of existing research. The *Canadian Climate Framing (CCF)* database was designed to address several of these gaps. Unlike earlier studies[3], the CCF database spans nearly five decades and draws on a larger set of newspapers, including not only nationally circulating outlets but also regional ones, thereby capturing a broader media landscape. It also better reflects Canada's linguistic diversity through its bilingual corpus, which includes a substantial share of French-language coverage (82.9% English, 17.1% French).

The CCF database is the first climate media corpus to offer sentence-level annotations derived from supervised machine learning. Unlike unsupervised approaches such as topic modeling, which remain dominant in the field[4], the methodology relies on transformer-based classifiers (BERT for English, CamemBERT for French) trained on over 4,000 manually expert-coded sentences. This supervised approach enables the identification of thematic frames, actor types, events, solution strategies, and emotional tone with a level of semantic precision that unsupervised methods cannot achieve. The trained models attain a macro F1 score of 0.866 across all 65 annotation categories, as validated against an independent gold standard of 1,000 expert-coded sentences whose reliability was confirmed through double-coding (Cohen's $\kappa = 0.569$, Krippendorff's $\alpha = 0.690$, Gwet's AC1 = 0.893). In addition, named entity recognition (NER) was also applied to extract persons, organizations, and locations mentioned in the articles, enabling, for example, network analyses of epistemic authorities and the mapping of actor co-citation patterns across the corpus.

Only two studies have applied comparable transformer-based models in this domain: Luo, Card, and Jurafsky [5] trained a single BERT classifier to analyze the evolution of partisanship in U.S. media, while Meddeb et al. [6] used a CamemBERT model to detect misinformation in French news. Importantly, neither of these studies examines framing directly. More importantly, no existing research has trained and deployed such a comprehensive suite of transformer models to systematically annotate climate-related media coverage. With over 60 BERT and CamemBERT classifiers covering frames, actors, events, solutions, and emotional tone, the CCF database fills a clear methodological gap by introducing a supervised NLP approach capable of detecting multiple analytical dimensions at scale, and test numerous hypotheses. The analytical applications presented in this paper illustrate the range of research questions this resource can support, and its capacity to inform studies across diverse theoretical perspectives.

This technical paper presents the methodology underlying the *Canadian Climate Framing (CCF)* database. With its fully pre-processed, analysis-ready structure, the CCF Database constitutes the most extensive resource to date for examining Canadian climate discourse. The remainder of the document outlines the full workflow, from data acquisition and preprocessing to model training and validation, and offers the conceptual rationale and practical steps necessary for researchers wishing to work with the CCF database or adapt its methodology elsewhere.

2 Methods

The CCF Database adopts a mixed-methods design that couples large-scale data collection with state-of-the-art machine learning (ML). It provides a systematic and reproducible framework for large-scale analysis of climate discourse that allows researchers to track how climate change is discussed, how narratives evolve over time, and who participates in the conversation.

Figure 1 presents the complete methodological pipeline and illustrates how the CCF Database transforms raw newspaper articles into a fully annotated database suitable for robust and reproducible climate communication research, both quantitative and qualitative. The pipeline integrates human expertise with computational methods through an iterative human-in-the-loop machine learning procedure. This approach ensured both the scalability needed to process 266,271 climate-related articles and the accuracy required for rigorous academic research.

The methodology comprises four distinct phases: (1) the *Collection Phase* involved systematic gathering of climate-related articles from 20 major Canadian newspapers using search queries resulting in 266,271 articles spanning 1978–2024; (2) the *Preprocessing and Design* phase transformed raw text into structured data, including sentence segmentation, language verification, and development of a comprehensive 65-category annotation framework; (3) the *Machine Learning* phase created and validated transformer-based machine learning models (BERT for English, CamemBERT for French) through iterative training with over 4,000 expert manual annotations; and (4) the *Validation and Deployment* phase ensured rigorous quality assurance by combining standard performance metrics with a human evaluation protocol based on an expert-coded gold standard that was independently double-coded by a second annotator to assess intercoder reliability, before deploying the trained models to annotate the complete corpus of 9.2 million sentences.

2.1 Phase 1 & 2: Data collection, preprocessing, database creation and annotation protocol

2.1.1 Step 1: Data collection

The CCF database was built through a systematic protocol to trace the evolution of climate discourse in Canada's bilingual print media, reaching back to the earliest available archives.

A total of 20 newspapers were selected for the analysis (see Appendix A, Table A2). The objective was to include both national and regional outlets in order to ensure broad geographic representation. Within this pool, newspapers were primarily chosen based on circulation levels [7], with the highest-circulating outlets retained. Additional criteria also informed the selection, including the desire to capture a range of political orientations and to build a bilingual corpus that reflects Canada's linguistic landscape. This included French-language outlets both inside and outside Quebec, as well as one English-language outlet located in Quebec[5]. The final selection was also partly shaped by availability constraints[6].

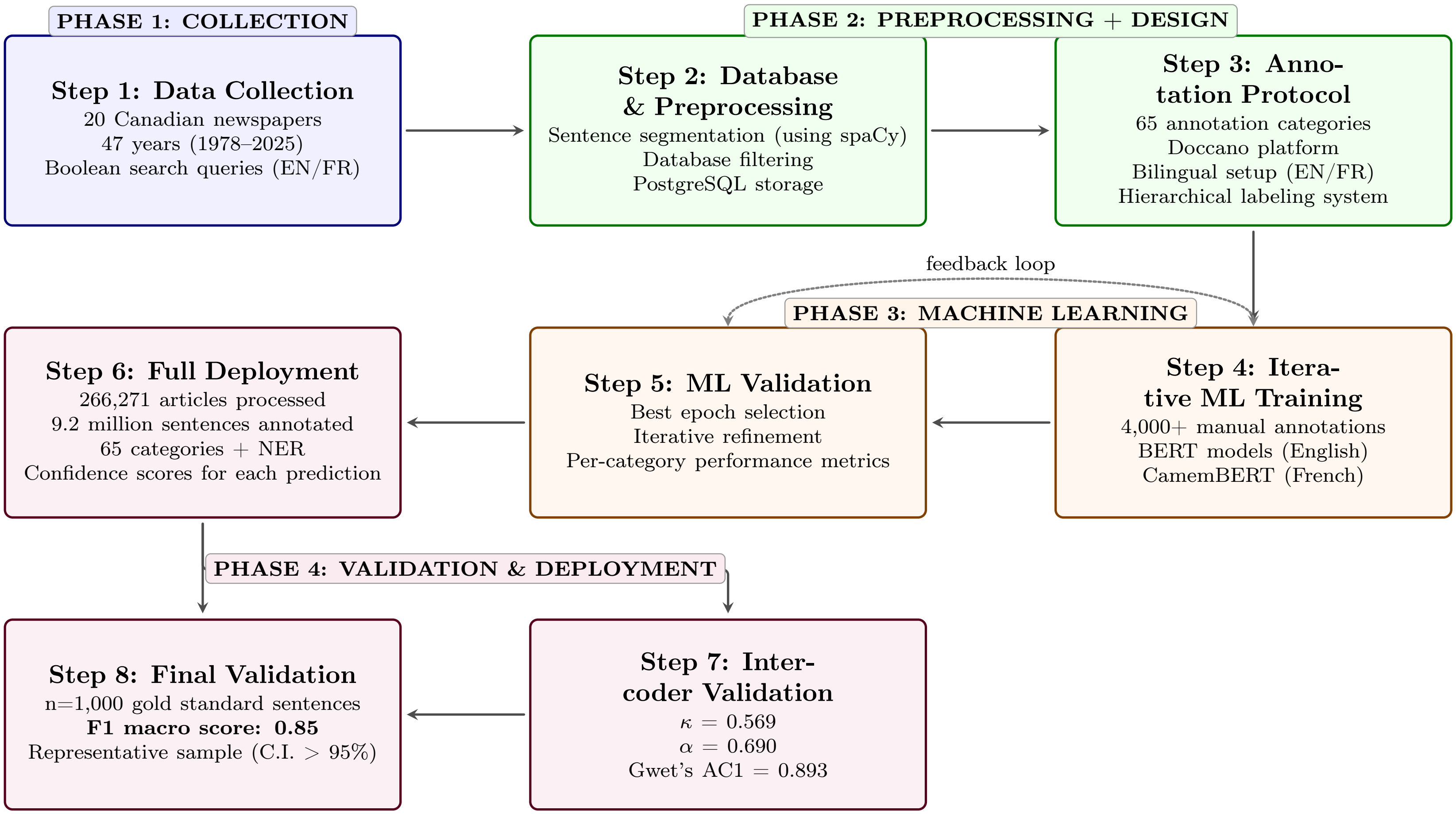

Figure 1: Methodological pipeline of the Canadian Climate Framing (CCF) project.

The corpus spans 47 years, from 1978 to 2024, with 1978 marking the earliest article identified. To identify climate-related articles, the authors developed comprehensive Boolean search queries tailored to each language (Table 1). The selected queries included a variety of terms related to the climate crisis. They were extensively tested beforehand to ensure optimal performance in capturing relevant articles. To be included in the data set, an article had to contain at least one of these keywords in its title or body. These queries yielded an initial corpus of over 300,000 articles.

Table 1. Boolean queries used for article retrieval

Language	**Boolean Query**
English	"global warming" OR "climate change" OR "climate disruption" OR "climate disturbance" OR "climate disturbances" OR "climate crisis" OR "greenhouse gas"
French	"réchauffement climatique" OR "réchauffement planétaire" OR "changement climatique" OR "changements climatiques" OR "dérèglement climatique" OR "crise climatique" OR "gaz à effet de serre"

2.1.2 Step 2: Preprocessing

The preprocessing phase transformed the raw data into a structured dataset suitable for machine learning. Each article was segmented into two-sentence contexts using language-specific spaCy models[7]. Sliding windows with single-sentence overlaps were implemented to ensure complete coverage while maintaining semantic continuity throughout each article. The choice to use two-sentence units was guided by several considerations. This level of granularity supports machine-learning applications by increasing the size of the training dataset. More importantly, working at the (near-)sentence level is the only approach that allows ex-post analysis that is highly precise and granular, because the unit of analysis is reduced to sentences rather than paragraphs, which in turn allows us to derive article-level variables by aggregating sentence-level annotations.

Quality control measures were applied systematically throughout the data preparation pipeline. Duplicate articles were identified and removed with fuzzy string-similarity algorithms set to a 95% threshold. Any article containing less than 100 words were also removed to eliminate reader letters, which often did not provide enough material to identify the presence of a frame and did not qualify as a press article per se. However, all other types of text (opinion columns, letters to the editor, news articles, editorials, etc.) were retained.[8]

Articles published by external news agencies as well (for example, *Agence France-Presse*, *CanWest News Service*, *Reuters*, *Bloomberg*) or by other newspapers not included in the analysis but which often belong to the same media conglomerate as the ones that were selected (for example, *the Financial Post*) were also retained. These republications align with the editorial direction of the concerned news outlet, and a republication does not contradict the objectives of the study since it indicates editorial interest in a topic concerning climate change. This process reduced the corpus to 266,271 unique articles[9]. Language was verified with fastText[10] to ensure accurate categorization of bilingual content, and dates were standardized and performed UTF-8 normalization. This data preparation produced a clean, standardized corpus ready for machine learning and annotation, which were stored in two

PostgreSQL tables: one with full metadata (including titles, main text, authors, dates, and page numbers) and another with the processed text data (including sentence segments and their corresponding IDs).

The resulting database is illustrated in Figure 2, which displays the distribution of article counts by media (20 outlets), while Figure 3 shows total article counts per year across the full time span untill 2024[11]. Finally, Figure 4 presents the distribution of articles by province, based on the primary location of each newspaper. This geographic breakdown shows the representativeness of the database across Canada (with 17.1% of French articles, and 82.9% of English articles). Three national newspapers are omitted from the provincial breakdown due to their pan-Canadian scope; nevertheless, they account for 36.2% of the articles: The Globe and Mail, the National Post, and the Toronto Star.

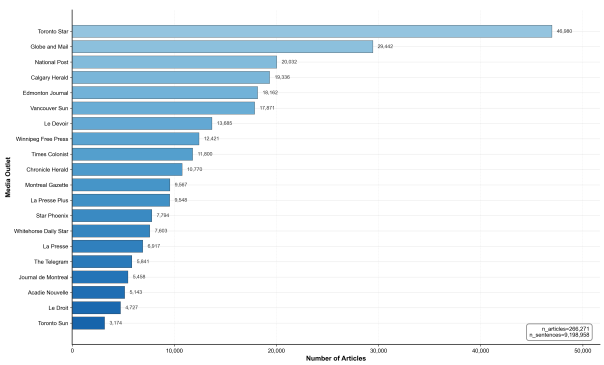

Figure 2: Distribution of articles by media.

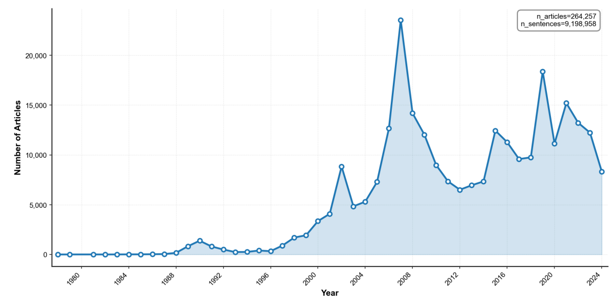

Figure 3: Total number of articles per year.

2.1.3 Step 3: Annotation protocol

The annotation framework constitutes the conceptual core of the CCF database. It comprises 65 categories organized into a hierarchical taxonomy designed to capture multiple dimensions of climate discourse at near-sentence granularity. The framework was developed through a combined theory-driven and data-driven procedure: it draws on established scholarship in framing, agenda-setting, climate communication, and social problems studies, while also being refined inductively through iterative engagement with the corpus and extensive pretesting. The full taxonomy and operational definitions are reported in Appendix B (Table B1), while Table 2 provides a compact overview of the framework's structure. Concretely, the framework is organized around *Primary Categories* that correspond to broad analytical dimensions, and *Sub-categories* that capture internal distinctions within each *Primary Categories* (Table 2).

Primary categories. The framework is organized around four primary analytic categories that structure the annotation protocol: *Thematic Frames*, *Actors/Messengers*, *Events*, and *Solutions* (Table 2). This design choice reflects a central insight of climate communication research: media discourse does not only describe climate change, but systematically selects interpretive lenses, attributes voice and credibility to specific actors, and connects the issue to episodic events and to (often contested) courses of action.

First, for *Thematic Frames*, we build on the general framing tradition often associated with Entman [8], which encompasses four dimensions: problem definition, causal interpretation, moral evaluation, and treatment recommendations. In practice, however, applying this

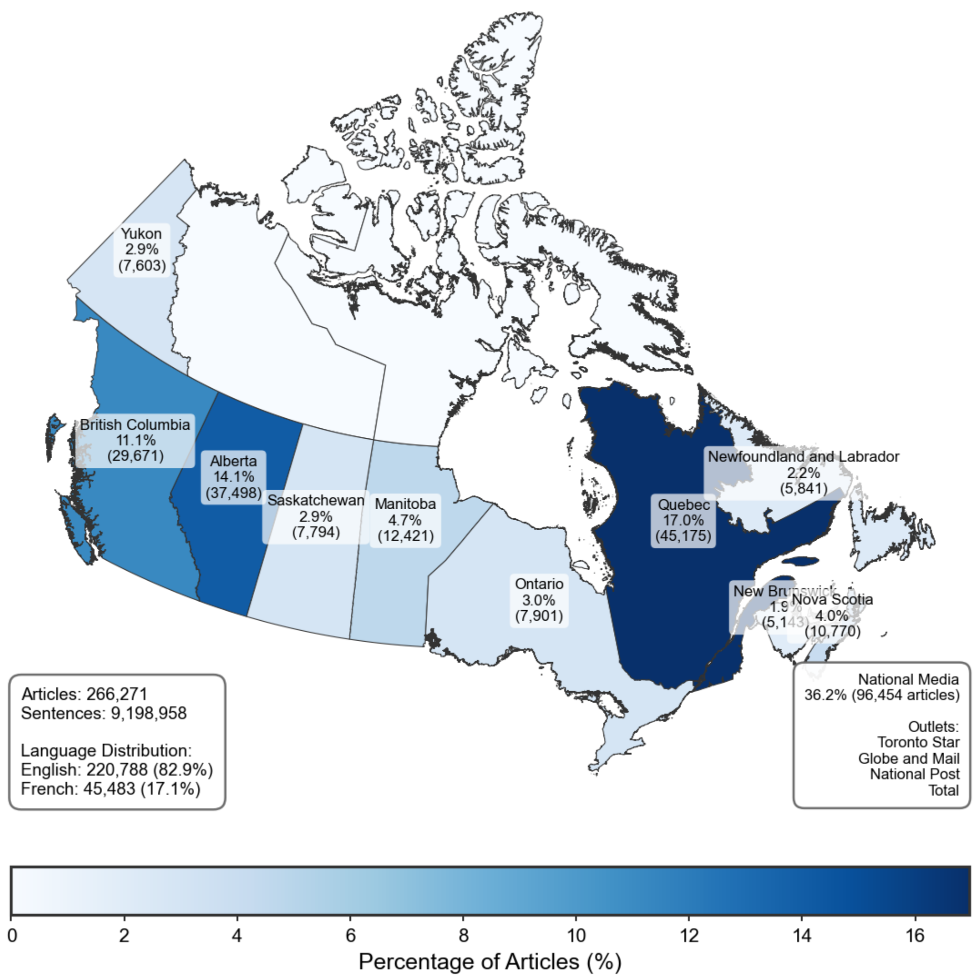

Figure 4: Distribution of articles by province.

framework to climate discourse presents significant challenges. Individual statements frequently span multiple dimensions, and "treatments" (i.e., proposed policy responses) prove particularly difficult to identify consistently [9]. This difficulty is amplified by the nature of climate change as a "wicked problem," where proposed solutions remain subject to scientific uncertainty and political contestation [10]. For these reasons, many studies do not apply Entman's four dimensions strictly [11]. The CCF framework addresses these challenges by operationalizing thematic frames around major societal domains in which climate change is repeatedly interpreted in media discourse: economic, security, public health, cultural, environmental, political, scientific, and justice spheres (Table 2). This domain-based approach retains core elements of problem definition, causal interpretation, moral evaluation, and treatment recommendations where they naturally emerge in the data.

Second, the framework treats *Solutions* as a distinct primary family rather than embedding them inside each thematic frame. This choice reflects both the empirical ambiguity of policy prescriptions in climate coverage and the substantive interest in distinguishing mitigation from adaptation, the latter having historically received less media attention [12]. Third, the *Actors/Messengers* family is grounded in the long-standing finding that *who* speaks about

Table 2: The CCF annotation framework: 65 hierarchical categories for comprehensive climate discourse analysis

Main Category	N[a]	Sub-categories and Detection Elements
Primary Categories: Thematic Frames (8 frames with 35 sub-categories)		
Economic	6	Negative/positive impacts, costs/benefits of action, sector footprint, general link
Health	5	Negative/positive impacts, co-benefits, healthcare footprint, general link
Security	5	Military response, base disruption, displacement, conflicts, defense footprint
Justice	5	Winners/losers, North-South responsibility, legitimacy, litigation, general link
Political	3	Policy measures, political debate & opinion, general link
Scientific	3	Controversy, discovery & innovation, general link
Environmental	3	Habitat loss, species loss, general biodiversity link
Cultural	5	Art, events, Indigenous practices, cultural footprint, general link
Primary Categories: Actors, Events and Solutions (with 15 sub-categories)		
Actors/Messengers	9	Scientists, politicians, activists, medical/economic/security/legal/cultural experts, any messenger
Events	6	Natural disasters, conferences, reports, elections, policy announcements, any event
Solutions	3	Mitigation, adaptation, any solution mention
Emotional Tone	3	Positive (hope, optimism), Negative (fear, anxiety), Neutral/none
Geographic Focus	1	Canadian places, actors, data, and policies
Urgency to act	1	Conveys immediate danger, crisis, or "code red" urgency
Named Entities	3	Persons (PER), Organizations (ORG), Locations (LOC)
Total: 65 categories		

[a]The N count includes detection of the main category plus its sub-categories. For example: Economic frame detection + 5 sub-categories = 6.

climate change matters for credibility, salience, and persuasion [13, 14]. Finally, the *Events* family reflects agenda-setting scholarship showing that media attention is cyclical and often driven by external shocks and institutional moments (e.g., disasters, major conferences, reports) [15, 16]. Together, these four primary categories provide a theory-informed structure for capturing how climate change is framed, who is authorized to speak, what triggers attention, and which actions are articulated (Table 2).

Operationalization: hierarchical taxonomy and sub-categories. Operationally, the framework implements a hierarchical labeling logic that mirrors its conceptual structure (Table 2). Each *Primary Category* contains (i) a detection-level category that captures whether the broader dimension is present in a two-sentence unit, and (ii) fine-grained sub-categories that specify which internal distinctions apply when the primary category is detected. This design enables multi-label annotation: multiple frames, actors, events, and solutions may co-occur within the same unit, and the same applies to sub-categories. Within *Thematic Frames*, each domain frame (e.g., Economic, Health, Justice) is paired with sub-categories that capture recurring interpretive distinctions observed in the literature and in the corpus (Table 2). Importantly, these sub-categories are not treated as mutually exclusive: a unit may simultaneously describe, for instance, the negative economic impacts of climate change and the economic benefits of climate action. The same principle applies to *Actors/Messengers*, where a single person may legitimately be coded as multiple expert types (e.g., a physician quoted about climate-health links may also be a natural scientist). For full category definitions, see Appendix B, Table B1.

Other dimensions: context, tone, urgency, and entity extraction. In addition to the four *Primary Categories*, the framework includes theoretically motivated dimensions that capture context and meaning-making beyond framing in the strict sense (Table 2). A *Geographic Focus* category records whether coverage situates climate change in a Canadian context, reflecting concerns that climate change is often portrayed as geographically distant [17]. An *Emotional Tone* category captures positive, negative, or neutral valence, consistent with research showing that emotions shape how audiences interpret and respond to climate discourse [18, 19]. An *Urgency to act* category captures alarmist or "code red" signals that communicate immediacy and crisis framing. Finally, Named Entity Recognition (NER) extracts and classifies *Persons*, *Organizations*, and *Locations* into a JSON field, enabling downstream analyses such as cocitation networks and the mapping of epistemic authorities (Table 2; Appendix B, Table B1).

Pretesting and iterative refinement. Finally, allongside these theoretical design choices, the framework was iteratively refined through inductive coding rounds. As sentences were annotated, emergent themes sometimes prompted targeted additions and clarifications to category definitions, followed by pretesting to ensure internal consistency and coder interpretability. This procedure strengthened the framework's construct validity and ensured that the final taxonomy remained both theoretically anchored and empirically adequate.The following real example from the database—an excerpt from a 2022 article—illustrates the corresponding manual annotation protocol:

Example (annotated sentence).

"In Scientific American, professor of Environmental Studies at Humboldt State University Sarah Jaquette Ray writes, 'Climate anxiety can operate like white fragility, sucking up all the oxygen in the room and devoting resources toward appeasing the dominant group.' But it would be a mistake to conclude that this term isn't applicable in the Global South, period."

Tagged Dimensions	Tagged categories
Actors/Messengers	Scientists
Thematic frames	Justice (primary category) + North–South responsibility; Health (primary category)
Emotional tone	Negative

2.2 Phase 3: Machine learning

The development of the training data followed an iterative manual annotation protocol aligned with the annotation framework previously described. A single expert annotator (climate policy and communication) labeled sentences in sequential batches (first 1,000 sentences, then three additional batches of 1,000 each) for a total of 4,000 annotated samples. The batching served two explicit objectives: (1) to iteratively improve and stabilize the annotation guidelines after each round; and (2) to progressively monitor and optimize model training to reach the highest attainable F1 score. This process culminated in a macro F1 score of 0.826 (see Table 3) during the training phase (see Table B2 in Appendix B for detailed metrics).

The sampling process drew equally from English and French language groups, with final annotation counts reaching 1,927 English and 2,073 French sentences. The training dataset was subsequently partitioned using stratified random sampling to create training (80%) and validation (20%) databases for each category and language. Table B3 in Appendix B provides the complete distribution of training and validation samples for all annotation categories. The random sampling approach ensured broad coverage across the temporal span and diverse media sources. The annotation protocol was developed with detailed operational definitions for each of the 65 categories to ensure consistency.

The machine learning pipeline leveraged transformer architectures optimized for each language through a custom fork of the AugmentedSocialScientist library [20, 21]. The fork built from Do, Ollion, and Shen [20] adds several functionnalities that are central to ensure robust machine learning training (metric logging at every training epoch, best-model selection using a weighted F1 score (0.7 × F1-positive + 0.3 × macro-F1), and an automated reinforced training protocol for underperforming models (described below) [21]). English text processing employed BERT-base-uncased, while French text utilized CamemBERT-base, both containing 110 million parameters. The training strategy implemented hierarchical classification, where detection models were trained first as binary classifiers on the full annotated dataset, followed by sub-classification models trained exclusively on manually annotated sentences that were positive for the corresponding primary category. This approach minimized false positive propagation and allowed for specialized optimization for each classification task.

Models failing to achieve positive-class F1 scores above 0.70 during the training phase underwent an automated reinforced training protocol[12]. The reinforced phase implemented weighted random sampling to oversample minority classes, doubled batch size to 64, reduced learning rate to 1e-5, and applied weighted cross-entropy loss with emphasis on the positive class. This reinforcement protocol, extending training for an additional 20 epochs, successfully improved mean F1 scores by 0.23 across affected models, and managed to bring most categories above the acceptable performance threshold. However, five categories could not be trained due to insufficient annotations in the training data and were consequently excluded from the annotation framework[13]

2.3 Phase 4: Deployment

The full deployment phase applied all the trained models to annotate the complete CCF Database of 9.2 million sentences across 266,271 articles. A hierarchical classification strategy was implemented that directly mirrors the annotation protocol's structure of primary categories and their associated sub-categories, as illustrated in Figure 5.

The hierarchical approach reflects the conceptual organization of the annotation framework with eleven primary detection categories, each with specific sub-categories. The eight thematic frames (*Economic*, *Health*, *Security*, *Justice*, *Political*, *Scientific*, *Environmental*, and *Cultural Frame*) function as primary categories with their internal distinctions—for example, *Economic Frame*, when positive (=1), triggers six economic sub-models (*nega-*

Table 3: Model training performance metrics for primary annotation categories

Category	**F1 (Class 1)**		**F1 (Class 0)**		**Macro F1**	
	EN	**FR**	**EN**	**FR**	**EN**	**FR**
Thematic Frames						
Economic Frame	0.745	0.814	0.944	0.957	0.845	0.885
Health Frame	0.800	0.667	0.989	0.994	0.894	0.830
Security Frame	0.870	0.800	0.996	0.997	0.933	0.898
Justice Frame	0.719	0.717	0.975	0.981	0.847	0.849
Political Frame	0.808	0.774	0.897	0.888	0.853	0.831
Scientific Frame	0.784	0.702	0.953	0.962	0.869	0.832
Environmental Frame	0.842	0.625	0.989	0.980	0.915	0.802
Cultural Frame	0.773	0.833	0.986	0.993	0.879	0.913
Other Primary Categories: Actors, Events and Solutions						
Presence of Messengers	0.912	0.929	0.904	0.915	0.908	0.922
Presence of Events	0.794	0.819	0.935	0.932	0.865	0.876
Presence of Solutions	0.737	0.878	0.914	0.944	0.825	0.911
Mention of Canada	0.942	0.964	0.968	0.980	0.955	0.972
Urgency to act	0.760	0.690	0.987	0.981	0.874	0.835
Overall Average*	**0.769**	**0.739**	**0.905**	**0.909**	**0.837**	**0.816**
Combined (ALL)	**0.754**		**0.907**		**0.826**	

*Overall average across all active annotation categories (see Table B2 for complete metrics).

tive/positive economic impacts of climate change, *economic disadvantages/benefits of climate action*, *carbon footprint of the economic sector*), while a negative detection (=0) bypasses these entirely.

Similarly, the three other primary categories follow this conditional logic: *Actors/Messengers Detection* activates nine actor-type sub-models (*Health Expert*, *Economic Expert*, *Security Expert*, *Legal Expert*, *Cultural Expert*, *Natural Scientist*, *Social Scientist*, *Activist*, *Public Official*), *Event Detection* triggers eight event-type classifications (*Extreme Weather Event*, *Meeting/Conference*, *Publication*, *Election*, *Policy Announcement*, *Judiciary Decision*, *Cultural Event*, *Protest*), and *Solutions Detection* applies two solution sub-categories (*Mitigation Strategy*, *Adaptation Strategy*). The deployment also processed standalone primary categories without hierarchical structure—*Emotional Tone* (*Positive Emotion/Negative Emotion/Neutral Emotion*), *Geographic Focus* (*Canadian Context*), *Urgency/Alarmism*.

In addition to the custom-trained classification models, pre-trained Named Entity Recognition (NER) models were employed, selected based on empirical evaluation for optimal performance on the dataset. A hybrid approach tailored to each language was implemented: for English texts, BERT-base-NER[14] was used for all three entity types (Person, Organization, Location), while for French texts, spaCy's fr_core_news_lg model[15] for person entities was combined with CamemBERT-NER[16] for organization and location entities. This hybrid strategy was used to leverage the respective strengths of each model: spaCy's superior performance on French person names and CamemBERT-NER's robust identification of French organizational and geographical entities (see Table B4 in the Appendix for performance metrics).

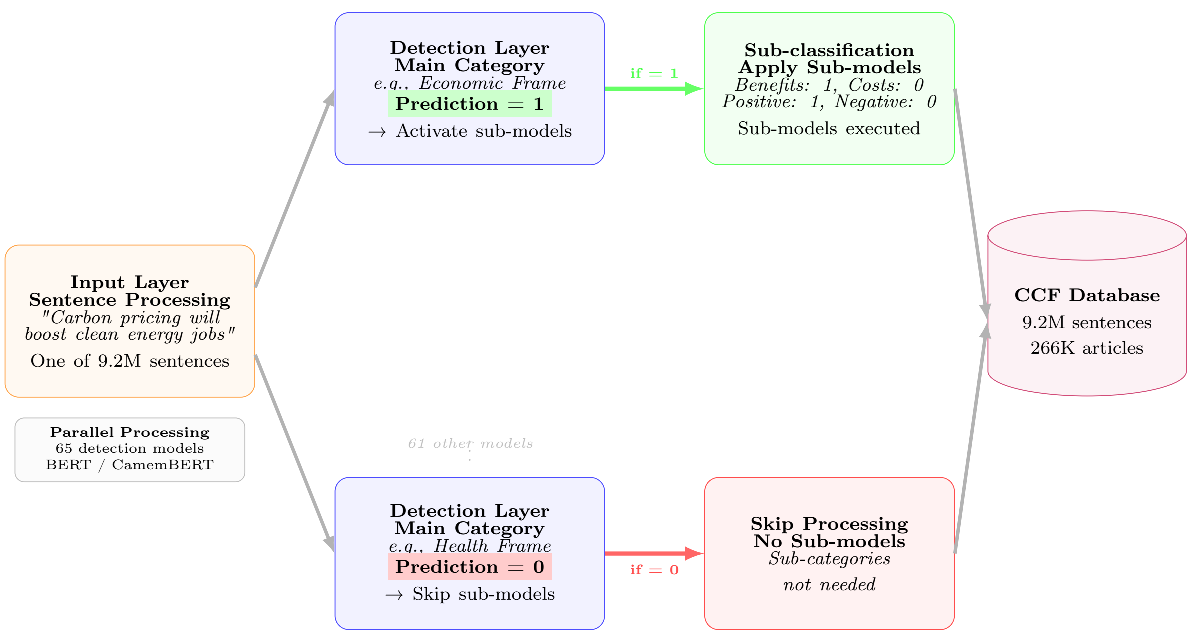

Figure 5: Hierarchical annotation pipeline. The system processes each sentence through detection models that determine whether corresponding sub-category models should be applied, enabling efficient annotation of the entire corpus.

3 Data Records

The CCF database is publicly available in a PostgreSQL-compatible format comprising two tables. The primary table (`CCF_processed_data`) contains approximately 9.2 million rows representing sentence-level analytical units with their annotations (binary variables), while a secondary table (`CCF_full_data`) stores article-level metadata for the 266,271 unique documents.

Due to copyright restrictions on the original newspaper content, raw sentence text is excluded from the public release. However, the depth of annotation, information extraction, and category multiplicity of the CCF Database was designed to produce high analytical value and enable comprehensive research (see Section 5.2 for example of analytical applications). Article-level metadata (titles, authors, publication dates, media outlets, page numbers), sentence-level identifiers, and all 65 machine-learning annotations are included. Researchers requiring access to the original sentence text for specific replication purposes may contact the authors directly.

The database structure follows the conceptual organization of the annotation framework illustrated in Figure 6: (1) *Metadata* columns (`doc_id`, `title`, `media`, `date`, etc.); (2) *Primary categories* for the eight thematic frames (`economic_frame`, `political_frame`, `scientific_frame`, etc.), actors/messengers (`messenger`), events (`event`), and solutions (`solution`); (3) *Sub-categories* providing fine-grained classifications (`eco_neg_impact`, `pol_debate`, `msg_scientist`, `evt_weather`, `sol_mitigation`, etc.); (4) *Other primary categories* including emotional tone (`tone_positive`, `tone_negative`, `tone_neutral`), geographic focus (`canada`), and urgency (`urgency`); and (5) *Named entities*. All annotation columns are stored as binary variables (1/0) indicating the presence or absence of each category, except for the `ner_entities` variable (*Named entities*) which contains extracted entity strings. All variable names are listed in the Code column of Table B1 (Appendix B).

The database is deposited at [repository to be specified] and includes: (1) the complete annotated database in PostgreSQL dump format; (2) CSV exports for compatibility with statistical software; (3) a data dictionary documenting all variables; (4) the complete expert-annotated training data (4,000+ manually coded sentences in JSONL format) and (5) intercoder reliability data from the validation protocol. The complete code for database construction, model training, and all analytical applications presented in this paper is available at `LINKOSF`.

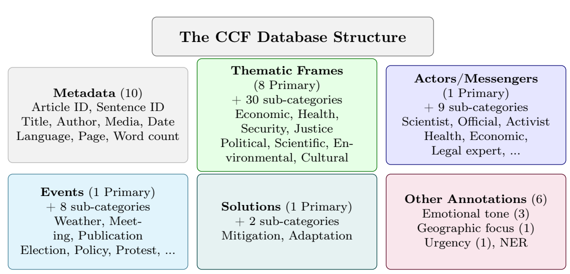

Figure 6: The CCF database structure organizes 76 columns into logical groups: metadata, primary categories (thematic frames, messengers, events, solutions), their sub-categories, and additional annotations. Sentence text is excluded for copyright compliance; all annotations and metadata are preserved.

4 Technical Validation

Technical validation of the CCF database and the trained models proceeded in three steps. First, a stratified validation set was constructed in line with established best practices for multi-label classification. Second, a PhD-level expert in climate change politics and policy produced the reference annotations that define the gold standard used for all model evaluation. Third, an independent second annotator replicated the coding following a structured protocol comprising a training phase (sentences 1–600) and a blind coding phase (sentences 601–1000), enabling assessment of inter-coder reliability.

4.1 Construction of the stratified validation set

The validation set comprises 1,000 sentences (500 per language) extracted from the fully annotated database using root-inverse probability weighting[17] combined with strict constraints to ensure statistical validity. Two constraints were applied: first, a minimum threshold of 40 positive examples per category to ensure sufficient statistical power for reliable performance estimation. Without this constraint, rare categories like *Disruption of military operations* (0.00004% prevalence in the database) would have too few examples to meaningfully assess model accuracy. Second, a 35% maximum allocation to prevent any single prevalent category from monopolizing the validation set. For instance, without this constraint, *Actors/Messengers Detection* (present in 48.8% of database sentences) could theoretically dominate the entire sample, leaving insufficient space for other categories.

4.2 Expert gold-standard annotation and model performance

All 1,000 sentences in the validation set were manually annotated by a PhD-level expert specializing in climate change politics and policy. These expert annotations define the gold standard against which all model performance metrics reported in this section are computed.

The validation results demonstrate robust performance across all annotation dimensions, with an overall F1 macro score of 0.866 (0.869 for English, 0.864 for French). As detailed

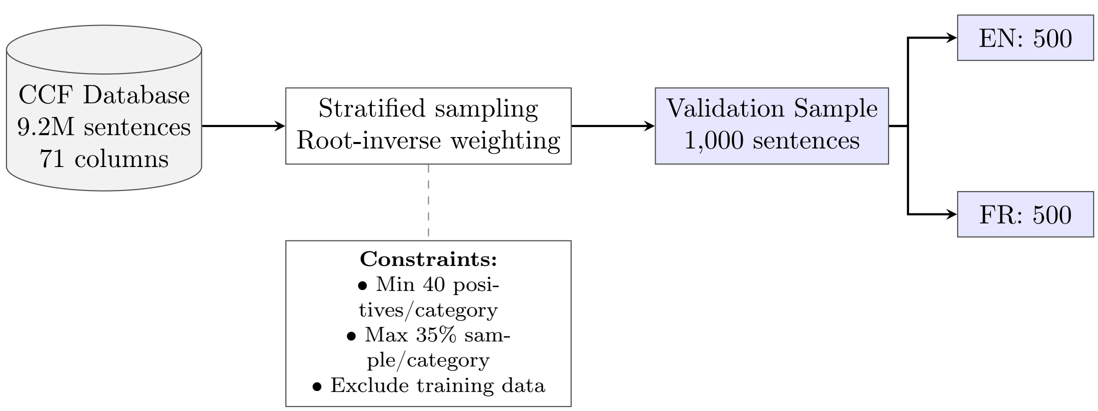

Figure 7: Stratified sampling procedure for final validation. Root-inverse probability weighting ensures balanced representation across all 65 annotation categories.

in Table B5 (Appendix), primary detection categories achieve the highest performance, with *Canadian Context* (F1=0.982) and *Actors/Messengers* (F1=0.961) showing near-perfect classification. Thematic frames maintain strong performance despite their semantic complexity, with *Scientific Frame* (F1=0.890) and *Environmental Frame* (F1=0.871) leading this category. The hierarchical classification strategy proves particularly effective, as evidenced by consistently higher recall scores for detection categories (mean recall=0.891) that trigger conditional sub-category evaluation. Notably, rare categories such as *Disruption of military operations* and *Post-disaster military assistance* show perfect precision despite minimal representation.

To assess temporal stability, stratified validation was conducted across five temporal periods (1980s–2020s) that revealed minimal performance drift (ΔF1 $<$ 1.1%) despite evolving media discourse patterns (Figure 8). This temporal consistency validates the models' generalization capacity across the full 46-year corpus span.

Table 4: Overall validation performance metrics

Language	**F1 Macro**	**F1 Micro**	**F1 Weighted**
English (EN)	0.869	0.909	0.911
French (FR)	0.864	0.902	0.905
Combined (ALL)	**0.866**	**0.905**	**0.908**

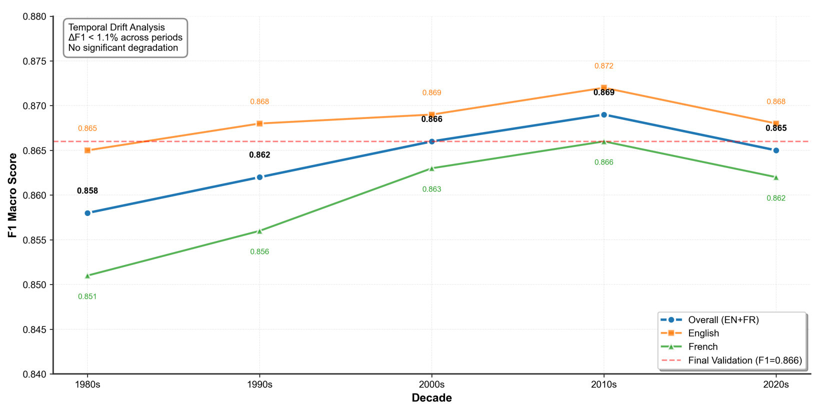

Figure 8: Temporal validation showing stable F1 macro scores across five time periods. The minimal variation (ΔF1 $<$ 1.1%) confirms robust model performance without temporal degradation.

4.3 Inter-coder assessment of the gold standard

Inter-coder reliability assessment was completed with a second independent annotator evaluating the same 1,000 randomly selected sentences from the validation set. The protocol followed a structured approach: during the first 600 sentences, the expert annotator provided training to the second coder, explaining category definitions and answering questions without directly intervening in the coding process. This training phase allowed the second coder to progressively learn the annotation framework while maintaining independence in coding decisions. After sentence 600, the annotation process became completely blind, with no communication between coders. The resulting comparison between the second annotator's labels and the expert gold-standard annotations provides an explicit evaluation of the reliability and reproducibility of the gold standard.

Given that the authors' main interest lies in the level of reliability once the codebook has been learned, the metrics for the blind coding phase (sentences 601–1000) are reported as the primary benchmark. Table 5 summarizes the agreement between the second annotator and the expert gold standard across all 65 annotation categories in this phase. These values indicate moderate agreement on Cohen's κ and substantial agreement on Gwet's AC1, while Krippendorff's α exceeds the commonly used 0.667 threshold for acceptable reliability. In other words, the gold-standard annotations produced by the expert can be reproduced to a reasonably high degree by an independent trained annotator, even in a setting with 65 multi-label categories and often dense and multi-topic climate discourse.

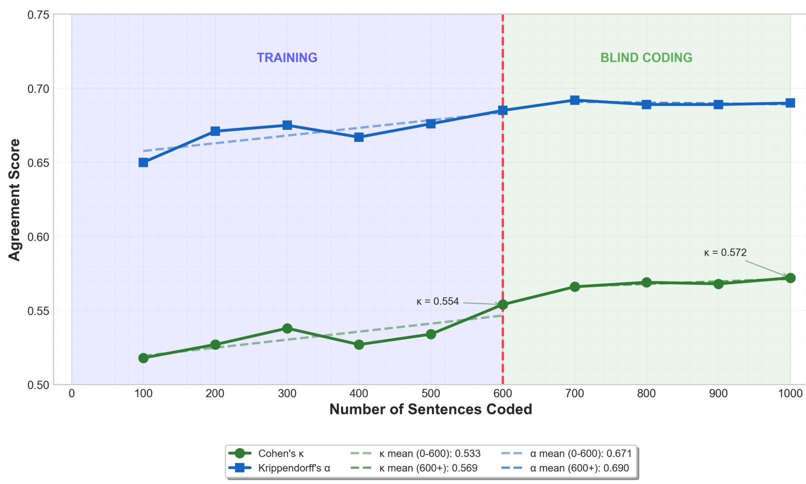

Figure 9: Inter-coder reliability progression across 1,000 annotated sentences, computed in successive blocks of 100 sentences.

However, several limitations should be kept in mind. The moderate κ values reflect the inherent difficulty of climate discourse annotation at the sentence level, where many sentences contain subtle framings and overlapping categories. Chance-corrected indices such as Cohen's κ are also known to be sensitive to category imbalance, which is pronounced in this setting. The substantially higher Gwet's AC1 suggests that part of the apparent disagreement in κ is driven by these prevalence-related paradoxes rather than by systematic confusion between coders.

To better understand how reliability evolves as the second annotator internalizes the

Table 5: Inter-coder reliability metrics for the blind coding phase (sentences 601–1000)

	Cohen's κ	Gwet's AC1	Krippendorff's α
Blind coding phase (601–1000)	0.569	0.893	0.690

codebook, agreement metrics in successive blocks of 100 sentences were computed. This analysis shows a pattern consistent with a learning curve for a complex codebook: both Cohen's κ and Krippendorff's α increase over time, with higher values in the later blocks than in the early ones. Figure 9 illustrates this progression across the 1,000 sentences, with a continuing improvement after the transition to blind coding at sentence 600. The upward trajectory of both indices across the 100-sentence blocks indicates that, once training is complete, the annotation framework can be applied consistently. This reinforces the view that the expert-defined gold standard is reproducible by trained coders.

5 Data Descriptors

This section demonstrates the analytical capabilities of the CCF database through two complementary perspectives. First, the key distributional characteristics of the annotated data are presented. These baseline distributions provide context for understanding the Canadian climate media landscape and identifying areas for research. Second, four illustrative applications demonstrate how the database's granular annotations can support original analyses: relationships between political actors and climate coverage, regional variations in scientific framing, frame based editorial prioritization, and the network structure of "epistemic authorities" (i.e., individuals cited as sources in climate discourse). These examples represent only a fraction of the database's analytical potential; additional statistical modeling incorporating external data sources and advanced computational methods is currently being developed.[18]

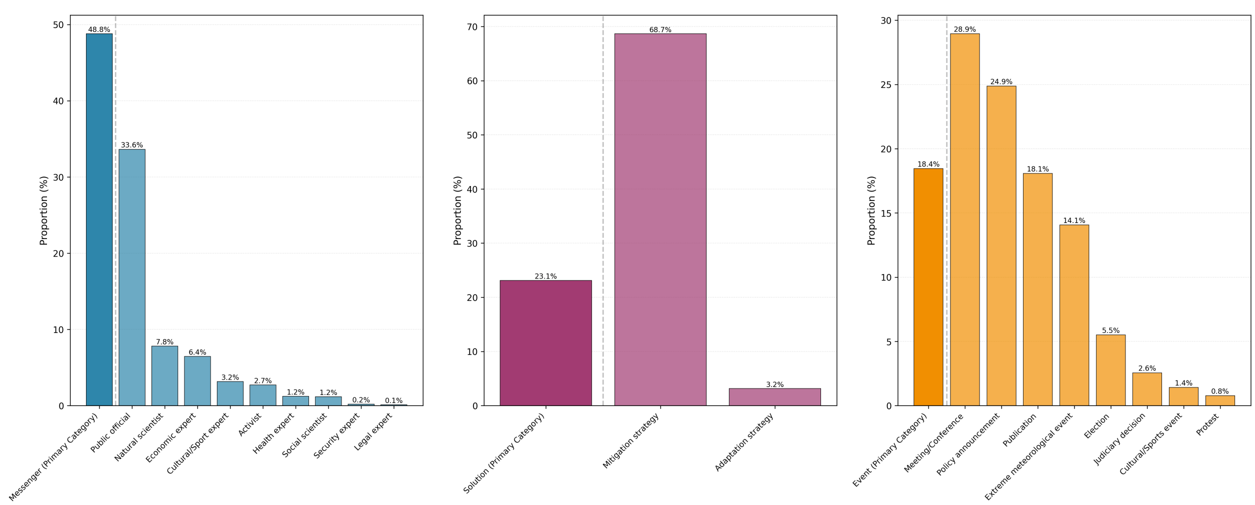

Figure 10: Distribution of annotation categories across the CCF database. The figure shows the average proportion of sentences containing each category for three major annotation groups: `messenger` (left), `solution` (center), and `event` (right). Primary categories represent sentences with any mention of the respective dimension, while subcategories provide more granular classification (see Table B1 in Appendix for detailed category definitions and variable codes).

5.1 Distribution of Annotation Categories Across the Database

To understand the composition and characteristics of the CCF database, Figure 10 presents the distribution of three key annotation groups across the database (see Table B1 in Appendix for all variable definitions). The `messenger` panel reveals that public officials (`msg_official`, 33.6%) and natural scientists (`msg_scientist`, 7.8%) are the most frequently cited actors in climate coverage, while economic experts (`msg_economic`, 6.5%) and cultural figures (`msg_cultural`, 3.2%) receive less attention. This distribution suggests that climate discourse in Canada remains primarily framed through political and scientific lenses. The `solution` panel demonstrates an overwhelming emphasis on mitigation strategies (`sol_mitigation`, 68.7%) compared to adaptation measures (`sol_adaptation`, 3.1%), indicating that Canadian media coverage focuses predominantly on reducing emissions rather than adaptation. The `event` panel shows that meetings and conferences (`evt_meeting`, 29.0%) and policy announcements (`evt_policy`, 24.9%) dominate event coverage, while protests (`evt_protest`, 0.8%) and cultural events (`evt_cultural`, 1.4%) receive minimal attention, suggesting an institutional bias in climate reporting.

The temporal dynamics presented in Figure 11 reveal strong shifts in climate discourse over the last five decades.[19] The most striking pattern is the rise of political framing (`political_frame`), which increased from virtually absent in the early 1980s to become the dominant frame by the mid 1990s, stabilizing around 35–40% of coverage in recent years. This politi-

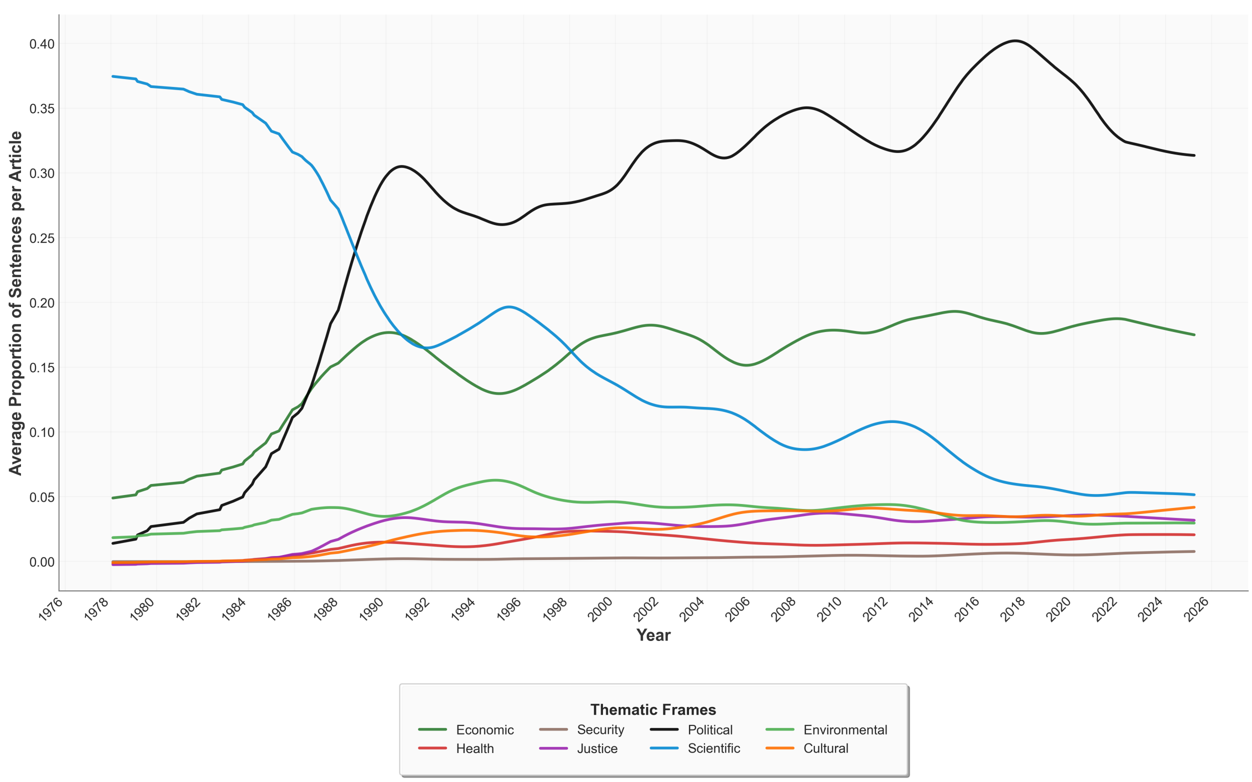

Figure 11: Temporal evolution of climate frames in Canadian media from 1978 to 2025. Lines represent LOESS-smoothed proportions of sentences containing each frame, with shaded areas indicating 95% confidence intervals calculated through bootstrap resampling (n=100).

cization corresponds with a dramatic decline in scientific framing (`scientific_frame`), which fell from approximately 40% in the late 1970s to less than 10% by 2025. The economic frame (`economic_frame`) started around 5% in the late 1970s, increased progressively to approximately 15% by the mid 1980s, and has since stabilized at this level. The environmental frame (`environmental_frame`), focusing on biodiversity and ecosystem impacts, has maintained a steady but modest presence around 3–5% of coverage. These patterns indicate a transformation from climate change as primarily a scientific issue, to one dominated by political and economical considerations, with implications for public understanding and policy development.

These distributional analyses demonstrate several key characteristics of the CCF database. First, the comprehensive coverage across all 65 annotation categories validates the framework's ability to capture the multidimensional nature of climate discourse. Second, the temporal and geographic patterns reveal how climate communication has evolved from scientific to political domains. Third, the dominance of institutional actors and events suggests opportunities for diversifying climate narratives to include more grassroots perspectives and adaptation strategies (which may reflect the actual reality of climate policies). Together, these patterns provide researchers with a baseline understanding of Canadian climate discourse that can inform more targeted analyses of specific frames, time periods, or regions.

5.2 Examples of Analytical Applications

While the descriptive statistics provide a foundation, the true potential of the CCF database lies in its capacity for accessible analyses, as data are already preprocessed, annotated and validated. Four scientifically grounded applications showcase this potential[20]: (1) how political actors shape climate coverage through their relationships with editorial prominence and thematic framing (variables `ner_entities`, and `sci_skepticism`); (2) a cartographic analysis of the geographic polarization of climate science discourse across provinces (variables `sci_skepticism`, and `media`); (3) how different thematic frames (`economic_frame`, `political_frame`, `scientific_frame`, `environmental_frame`, `justice_frame`, `cultural_frame`) influence newspapers' decisions about front page placement; and (4) the network structure of epistemic authorities, identified through Named Entity Recognition (variable `ner_entities`). These are the individuals cited in climate coverage whose cocitation patterns reveal the underlying social architecture of climate communication in Canadian media.

5.2.1 Political Actors in Climate Coverage

The CCF database supports targeted analysis of how specific political actors shape climate discourse. Through Named Entity Recognition (`ner_entities` variable) applied to all sentences, the database extracts person, organization, and location entities, allowing researchers to identify who speaks about climate change and how they are discussed. Figure 12 presents the 50 most frequently mentioned persons in 2024 climate coverage.

The results show that climate coverage in 2024 was dominated by political figures rather than scientists or climate experts. Justin Trudeau leads with 11,433 mentions (5.7% of all person mentions), followed by Donald Trump (9,206, 4.6%) and Pierre Poilievre (6,151, 3.1%). Federal ministers (Steven Guilbeault, Chrystia Freeland), provincial premiers (Danielle Smith,

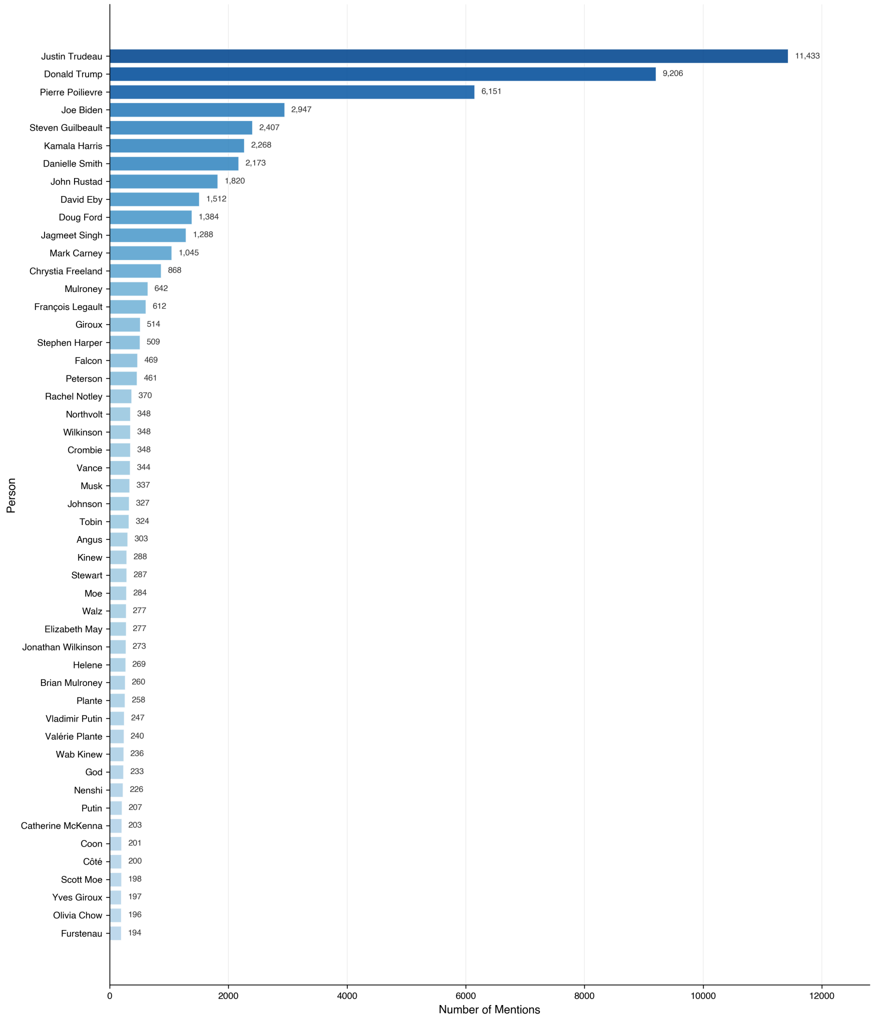

Figure 12: Top 50 most frequently mentioned persons in Canadian climate coverage (2024). Person entities are extracted from the `ner_entities` JSON field (PER category) across all 324,010 sentences. Names are normalized to consolidate variants (e.g., "Trump" and "Donald Trump" are merged). The dominance of political figures, with Justin Trudeau (11,433 mentions), Donald Trump (9,206), and Pierre Poilievre (6,151) at the top, reflects the politicization of climate discourse in Canadian media.

Doug Ford, David Eby), and U.S. political figures (Joe Biden, Kamala Harris) round out the top mentions. This pattern suggests that Canadian media frames climate change primarily through a political lens, with elected officials serving as the dominant voices in climate discourse.

Focused analysis on political leaders. Beyond identifying who appears in climate coverage, the CCF database allows examination of how specific actors relate to the content of that coverage. Figure 13 examines whether articles mentioning Canada's two major federal party leaders, former Prime Minister Justin Trudeau and Conservative Leader Pierre Poilievre, exhibit different patterns in scientific skepticism framing. Using separate univari-

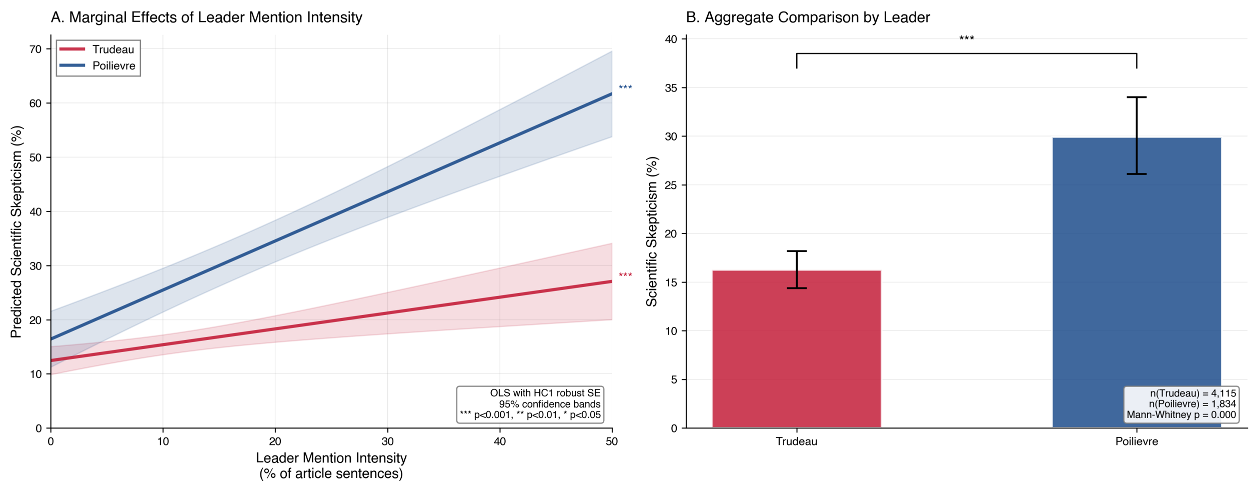

Figure 13: Effect of political leader mentions on scientific skepticism (`sci_skepticism`) from September 2022 to December 2024. Panel A shows marginal effects from OLS regression with 95% confidence bands; Panel B shows aggregate proportions with 95% bootstrap confidence intervals.[23]

ate OLS regressions with heteroscedasticity-robust standard errors (HC1), the relationship between leader mention intensity (proportion of article sentences mentioning each leader) and the `sci_skepticism` variable is modeled for each leader independently.[21]

The results reveal a strong asymmetry. Panel A shows that increased mention intensity of Pierre Poilievre is strongly associated with higher scientific skepticism ($\beta = +0.91$, $p < 0.001$), while Justin Trudeau shows a weaker but still significant association ($\beta = +0.29$, $p < 0.001$). The effect of Poilievre on skepticism is approximately three times stronger than that of Trudeau. Panel B confirms these patterns at the aggregate level: articles mentioning Poilievre contain significantly higher proportions of scientific skepticism classified sentences compared to articles mentioning Trudeau. Based on 25,351 articles from the period when Poilievre served as Conservative leader, 4,115 articles (16.2%) mention Trudeau and 1,834 (7.2%) mention Poilievre, with 1,331 (5.3%) mentioning both leaders. These findings suggest that media coverage associating Poilievre with climate issues more frequently employs frames that question scientific consensus, potentially reflecting both the leader's public stance on climate policy and editorial choices about how to contextualize his positions. This type of actor frame association analysis, impossible with traditional content analysis methods, becomes straightforward with the CCF database's preannotated structure.

5.2.2 Geographic Polarization of Climate Science Discourse

The CCF database also supports detailed geographic analysis of how climate science is discussed across Canadian provinces. Figure 14 reveals regional patterns in scientific skepticism. The analysis focuses on the `sci_skepticism` subcategory (see Table B1 in Appendix), which captures sentences that question the validity, accuracy, or sufficiency of climate science, including challenges to climate models, dismissals of scientific evidence, or expressions of doubt about the scientific consensus.

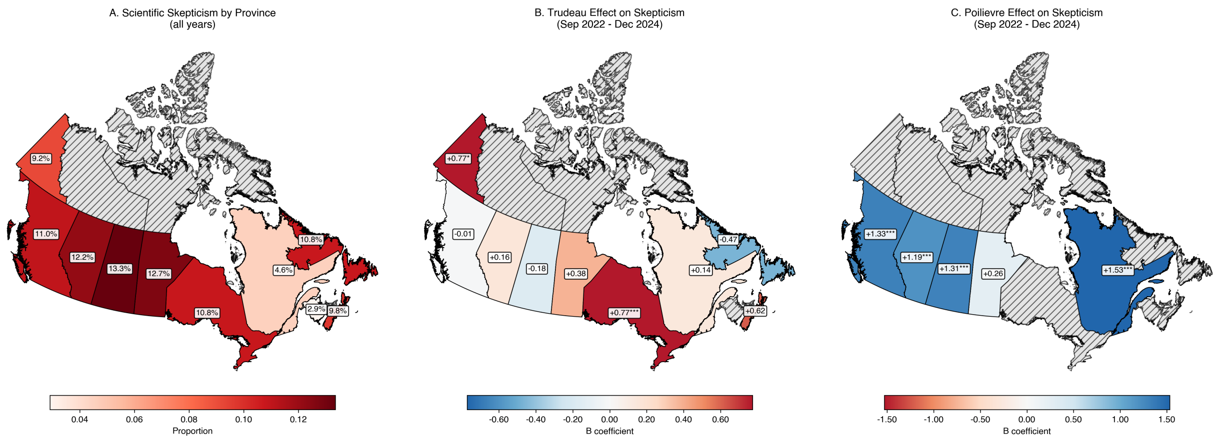

Figure 14: Geographic patterns of scientific skepticism (`sci_skepticism`) in Canadian climate coverage. Panel A shows the proportion of scientific skepticism among scientific sentences (all years). Panels B and C show the effect of leader mention intensity on scientific skepticism framing (September 2022–December 2024), estimated via OLS regression with robust standard errors.[26]

Panel A reveals geographic variation in the prevalence of scientific skepticism across provinces. Among sentences with scientific content, the proportion expressing skepticism varies notably across regions. Western provinces show distinct patterns compared to Eastern Canada. The Prairie provinces (Saskatchewan, Alberta) along with British Columbia display varying rates of skepticism framing, while Quebec and the Atlantic provinces show different levels.

Panels B and C extend this geographic analysis by examining how political leader mentions correlate with scientific skepticism framing at the provincial level, using separate univariate OLS regressions for each province-leader combination.[24] The contrast between leaders is striking: Trudeau mentions show a significant positive association with skepticism in Ontario ($\beta = +0.77$, $p < 0.001$) and Yukon ($\beta = +0.77$, $p < 0.05$), with no significant effects in other provinces. In contrast, Poilievre mentions are significantly associated with increased skepticism across multiple provinces: Quebec ($\beta = +1.53$, $p < 0.001$), British Columbia ($\beta = +1.33$, $p < 0.001$), Saskatchewan ($\beta = +1.31$, $p < 0.001$), and Alberta ($\beta = +1.19$, $p < 0.001$). This asymmetry suggests that coverage mentioning the Conservative leader is more systematically linked to skeptical framings of climate science, particularly in Western provinces and Quebec.

5.2.3 Frame Based Editorial Prioritization

The CCF database can reveal patterns in how different climate frames influence editorial decisions about story prominence. Rather than simply comparing frame presence, this analysis models the relationship between frame intensity (the proportion of article sentences containing each frame) and front page placement probability using separate univariate logistic regressions for each frame.[27]

Figure 15 shows a striking pattern in editorial prioritization. Panel A displays predicted front page probability curves as frame intensity increases from 0% to 100% of article con-

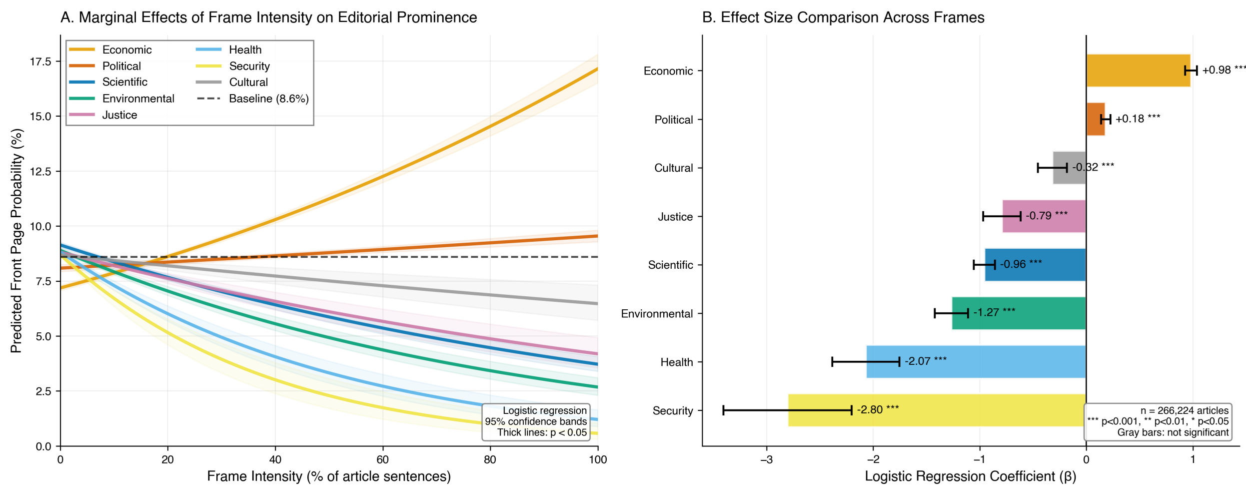

Figure 15: Effect of frame intensity on front page placement probability. Panel A shows predicted probabilities from logistic regression as frame intensity increases; Panel B shows regression coefficients with 95% confidence intervals.[29]

tent. Only two frames show positive associations with editorial prominence: `economic_frame` ($\beta = +0.98$, $p < 0.001$) and `political_frame` ($\beta = +0.18$, $p < 0.001$). As the proportion of economic framing increases in an article, front page probability rises substantially, from the 8.6% baseline to over 20% for articles dominated by economic content. All other frames exhibit significant negative associations: `security_frame` ($\beta = -2.80$), `health_frame` ($\beta = -2.07$), `environmental_frame` ($\beta = -1.27$), `scientific_frame` ($\beta = -0.96$), `justice_frame` ($\beta = -0.79$), and `cultural_frame` ($\beta = -0.32$), all $p < 0.001$. Panel B visualizes these effect sizes, with positive coefficients (right of zero) indicating frames that increase editorial prominence. These findings suggest that Canadian newspapers systematically privilege economic and political framings of climate change in their most prominent coverage, while scientific, environmental, and health perspectives receive lower editorial priority. This pattern has potential implications for public understanding of climate change as primarily an economic or political issue rather than an environmental or scientific one.

5.2.4 Network Structure of Epistemic Authorities

The CCF database also supports network analysis of "epistemic authorities," the individuals cited in climate discourse. Using Named Entity Recognition (`ner_entities`), this approach extracts all persons mentioned in climate coverage and maps their cocitation patterns (i.e., cooccurrence within the same article), thus revealing the underlying social structure of climate communication. Further analyses can filter by messenger type using the `messenger` category and its subcategories (`msg_scientist`, `msg_official`, `msg_activist`, etc.) to focus specifically on quoted sources.

Figure 16 presents a cocitation network of persons mentioned in 2024 climate coverage.[30] The 2024 network topology shows a highly centralized structure dominated by political figures. Justin Trudeau (1,485 citations), Pierre Poilievre (933 citations), and Donald Trump (793 citations) form the central core, while scientific authorities occupy peripheral positions.

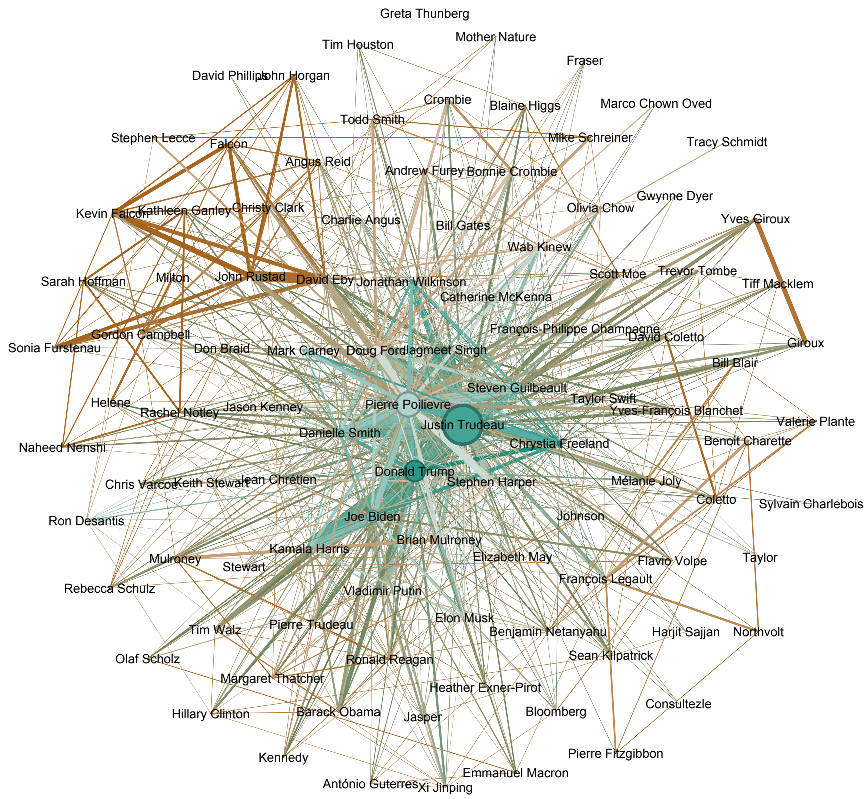

Figure 16: Cocitation network of epistemic authorities in Canadian climate media discourse (year=2024). Node size represents degree centrality, edge thickness indicates cocitation frequency, and colors denote community detection clusters. The network reveals a hierarchical structure with political figures at the core (Trudeau: degree centrality 0.899, Poilievre: 0.818, Trump: 0.667) and six interconnected communities: federal Canadian politics (community 3), US/international figures (community 1), Ontario (community 6), Québec (community 0), British Columbia (community 2), and historical political figures (community 5). Activist Greta Thunberg (community 4) appears isolated with no cocitation links to other top authorities.

The network exhibits high clustering (coefficient = 0.686), with seven detected communities: a dominant federal Canadian cluster (Trudeau, Poilievre, Guilbeault, Freeland, Wilkinson, Singh), a US/international cluster (Trump, Biden, Harris, Putin, Musk), provincial clusters for Ontario (Ford, Chow, Crombie), Québec (Legault, Charette, Plante), and British Columbia (Eby, Rustad, Furstenau), a historical figures cluster (Mulroney, Reagan, Thatcher, Pierre Trudeau, Chrétien), and an isolated community containing only Greta Thunberg, the sole activist with no cocitation links to other top 100 authorities.

The dominance of political over scientific authorities is clear: 19 of the top 20 authorities are politicians, Elon Musk (rank 18) is the only nongovernmental figure (at least at the time). This pattern reveals that Canadian media frames climate change primarily through political rather than scientific lenses. Combined with the earlier finding that articles mentioning Poilievre are significantly more associated with scientific skepticism framing (`sci_` `skepticism`), this network structure suggests that expertise itself becomes politicized in climate communication. The high clustering coefficient indicates that media coverage reinforces topical silos rather than fostering dialogue across domains, with authorities cited within specialized communities that rarely intersect.

6 Bridging Scale and Granularity in Climate Discourse Analysis

The Canadian Climate Framing (CCF) database establishes a new paradigm for climate communication research by combining scale, granularity, and machine learning within a single infrastructure. Transforming 266,271 articles across 46 years into 9.2 million analysis ready sentences with 65 validated annotations variables and NER, the CCF supports new analyses that were impossible before. Its four phase architecture and reproducible annotation framework address longstanding methodological limitations and can be adapted to examine contexts beyond Canada, supporting comparative work on how media agendas interact across different national settings.

With an overall F1 score of 0.866, and strong performance even for underrepresented categories, the database shows that computational methods can capture the semantic complexity of climate discourse with sufficient accuracy for rigorous research. Although the analytical categories and newspaper selection inevitably involve interpretive choices and remain open to refinement, the CCF database nonetheless represents a significant advance.

Beyond its technical contribution, the CCF provides a dynamic research infrastructure: it will be regularly updated to incorporate the latest coverage, and scholars can experiment with its annotation categories to explore frame combinations, messenger strategies, emotional tone, or spatiotemporal patterns. Ultimately, the CCF database helps to strengthen climate communication as an evidence-informed social science and provides a foundation for developing more effective communication strategies.

7 Author Contributions

Author 1: Designed and implemented the complete data architecture, including data extraction, cleaning, parsing, and database construction. Developed all machine learning models and methodological approaches. Wrote all code and conducted quantitative analyses. Drafted the methodology and results sections.
Author 2: Developed the theoretical and conceptual foundations underlying the annotation framework. Designed the complete annotation scheme and performed all manual annotation. Drafted the theoretical background and literature review sections.
Author 3: Contributed to conceptual development and analysis design. Verified statistical analyses. Reviewed and edited the complete manuscript.

8 Acknowledgements

The authors thank Richard Nadeau for his valuable input throughout the development of this project. The authors also acknowledge the research assistance of Maëlle Rouyer. We are also grateful to the Centre interuniversitaire de recherche sur la science et la technologie (CIRST) for its valuable support and guidance, including access to its research infrastructure. Earlier versions of this paper were presented at the annual meetings of the Canadian Political Science Association, the Prairie Political Science Association, and the American Political Science Association, as well as at the University of Montreal. The authors wish to acknowledge the anonymous reviewers and editor for their feedback throughout the development of this paper.

9 Funding

Funding for this research was provided by the Fonds de recherche du Québec [FRQ Grant number 337081].

Notes

1. While digital and social media have grown in prominence over the years, newspapers remain a foundational source of information and continue to shape public discourse. Their coverage also influences other media formats [22, 23]. Focusing on print media therefore provides a vantage point on a traditional and still highly influential channel of climate communication. Crucially, analyzing newspapers allows for long-term historical comparison, including periods that predate the rise of social media.

2. Canada offers a particularly compelling case for studying climate discourse. Its federal structure, bilingual culture, and vast territory create a microcosm of the diversity, and contradictions, found globally. While Canada is highly vulnerable to the impacts of climate change [24], it remains significantly dependent on fossil fuel extraction and other resource-based industries [25]. In turn, this context creates structural tensions between environmental protection and economic interests. Some of Canada's provinces, notably Quebec and British Columbia tend to adopt more progressive climate policies, whereas others with resource-intensive economies are more cautious or resistant [26]. This combination of political decentralization, linguistic duality, and regional economic variation results in diverse policy approaches and consequently shapes how climate change is discussed in the media.

3. Only a small number of studies have examined media coverage of climate change in the Canadian context, and those that do generally rely on a narrow set of nationally circulating newspapers—most commonly *The Globe and Mail*, *National Post*, and *Toronto Star*. Yet, climate attitudes vary significantly across regions [27], suggesting that media coverage may also differ geographically. Nevertheless, very few studies have investigated regional climate coverage, and those that do often focus on a single province or territory rather than comparing across them. This national-level bias also produces a linguistic blind spot: in an officially bilingual country, reliance on the most widely circulated papers effectively excludes French-language media. Only two studies have examined francophone coverage in Canada—[28], comparing English and French climate change media coverage, and [29], analyzing climate change's public-health framing with partial inclusion of Quebec. A further limitation is temporal scope. Many studies examine short periods, frequently a single year, while longitudinal analyses are rare and typically cover only the late 1980s to the 2010s. As a result, most corpora remain small—often fewer than 10,000 articles—and provide limited insight into how climate discourse evolves across regions, languages, and time.

4. Natural language processing (NLP) enables scholars to examine much larger corpora in far less time, opening new avenues for systematic inquiry [30, 31, 32]. However, as summarized in Table 2, when climate researchers do employ NLP, it is most often to identify key actors through named entity recognition, detect stance, or conduct sentiment analysis. These applications tend only to map actors, tone, or evaluative direction rather than interpretive structures. The dominant computational approach in the field continues to be unsupervised machine learning, particularly topic modeling, where topics are treated as proxies for frames. However, as Nicholls and Culpepper [33] and Otmakhova, Khanehzar, and Frermann [34] emphasize, this assumption is problematic: topics are too narrow to capture the defining features of frames, which, following Entman [8], require not only identifying a problem but also specifying causal interpretations, moral evaluations, and possible solutions.

5. Quebec is the only province in Canada with French as its sole official language. However, New Brunswick is officially bilingual, and several other provinces, such as northeastern Ontario along the Quebec border, also have sizeable francophone communities.

6. For instance, when the highest-circulating newspapers were unavailable, the next highest-circulating outlets were selected. In the cases of Prince Edward Island, Nunavut and the Northwest Territories, no eligible newspapers were available for inclusion.

7. en_core_web_lg was employed for English text and fr_dep_news_trf for French text, both optimized for news content processing.

8. Regarding this matter, Lawrence [35] explains that it is precisely in editorials, or more generally in opinion texts, where journalistic standards are applied with a greater degree of flexibility, that the battle of framing is best observed.

9. Duplicates of the same article from different media outlets were retained to enable the measurement of media bubbles.

10. fastText is an open-source library by Facebook AI Research for text classification and word representations; its pre-trained language identification model detects the language of a text with high accuracy. See `https://fasttext.cc/`.

11. Data extraction concluded in February 2025; consequently, 2025 was excluded from the yearly counts because the year is incomplete. Nevertheless, the authors plan to update the database regularly and create an observatory.

12. This protocol was necessary for 15 underperforming models, primarily in abstract conceptual categories such as cultural and justice sub-frames.

13. *carbon footprint of the health sector*, *positive impacts of climate change on health*, *post-disaster military assistance*, *Disruption of military operations due to climate change*, and *carbon footprint of the defense sector*.

14. Available at: `https://huggingface.co/dslim/bert-base-NER`

15. Available at: `https://spacy.io/models/fr`

16. Available at: `https://huggingface.co/Jean-Baptiste/camembert-ner`

17. The weighting formula $w_i \propto \sum_j 1/\sqrt{f_j}$ assigns higher sampling probability to sentences containing rare categories, where f_j represents the frequency of category j present in sentence i. For example, a sentence annotated with both *Economic Frame* (appearing in 15.4% of the corpus) and *Loss of Indigenous Practices* (appearing in 0.28% of the corpus) would receive a weight of approximately $1/\sqrt{0.154} + 1/\sqrt{0.0028} = 2.55 + 18.90 = 21.45$, making it roughly 8 times more likely to be selected than a sentence containing only *Economic Frame*.

18. For instance, the authors are developing media cascade detection algorithms that leverage the full cross referenced metadata (journalists, outlets, temporal patterns) to identify information diffusion dynamics across the Canadian media ecosystem. See `https://github.com/antoinelemor/CCF-media-cascade-detection` for ongoing work in this direction.

19. All proportions in this analysis represent the average proportion of sentences per article containing each frame, calculated at the article level before temporal or geographic aggregation.

20. All analysis scripts producing the figures and statistics in this section are available at `LINK`.

21. For each leader $\ell \in \{\text{Trudeau}, \text{Poilievre}\}$, we estimate: $\texttt{sci\_skepticism}_i = \alpha + \beta_\ell \cdot \text{intensity}_{\ell,i} + \varepsilon_i$, where $\texttt{sci\_skepticism}_i$ is the proportion of sentences questioning climate science in article i, and $\text{intensity}_{\ell,i}$ is the proportion of sentences mentioning leader ℓ. The variable `sci_skepticism` captures sentences that question the validity, accuracy, or sufficiency of climate science (see Table B1 in Appendix). Models use HC1 robust standard errors. Leader mentions are identified through pattern matching on the `ner_entities` JSON field (PER category). Panel B compares articles mentioning each leader at least once; Mann-Whitney U tests assess between-group differences. Analysis covers September 10, 2022 (when Poilievre became Conservative leader) to December 31, 2024.

22. For methodology, see Note 21.

23. For methodology, see Note 21.

24. For each province p and leader $\ell \in \{\text{Trudeau}, \text{Poilievre}\}$, we estimate: $\texttt{sci\_skepticism}_i = \alpha_p + \beta_{p,\ell} \cdot \text{intensity}_{i,\ell} + \varepsilon_i$, where $\text{intensity}_{i,\ell}$ is the proportion of article sentences mentioning leader ℓ. The variable `sci_skepticism` captures sentences that question the validity, accuracy, or sufficiency of climate science (see Table B1 in Appendix). Coefficients β represent the change in skepticism framing per unit increase in leader mention intensity. Analysis covers September 10, 2022 (when Poilievre became Conservative leader)

to December 31, 2024. Only provinces with $\geq$50 articles and $\geq$10 articles mentioning the leader are shown. Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Hatched areas indicate insufficient data.

25. For methodology, see Note 24.

26. For methodology, see Note 24.

27. For each thematic frame $f \in \{$`economic_frame`, `political_frame`, `scientific_frame`, `environmental_frame`, `justice_frame`, `health_frame`, `security_frame`, `cultural_frame`$\}$, we estimate a separate univariate logistic regression model: $\text{logit}(P(\text{FrontPage})) = \alpha_f + \beta_f \cdot \text{intensity}_f$, where intensity_f is the proportion of article sentences containing frame f. Coefficients β_f represent the change in log-odds of front page placement per unit increase in frame intensity. Predicted probabilities in Panel A are computed as $\hat{P} = 1/(1 + e^{-(\hat{\alpha}_f + \hat{\beta}_f \cdot x)})$. Confidence bands use the delta method. All models fit on 266,224 articles (22,884 front page, baseline rate: 8.6%). All coefficients significant at $p < 0.001$.

28. For methodology, see Note 27.

29. For methodology, see Note 27.

30. The network analysis proceeds in four stages: (1) person entities are extracted from the `ner_entities` field across all 2024 sentences; (2) names are normalized using automated resolution and manual verification for high-profile figures; (3) persons are aggregated at the article level; (4) cocitation networks are constructed where an edge connects two persons mentioned in the same article. For visualization clarity, only the top 100 persons by mention frequency are shown, with edges requiring at least 3 cocitations. For more details, see https://github.com/antoinelemor/CCF-canadian-climate-framing.

A Appendix A: Corpus Composition and Previous Studies

Table A1: Summary of previous studies analyzing climate change coverage in Canadian newspapers

Study	Newspaper(s)	Scope	Language	Time Period
Good [36]	The Globe and Mail	National	English	2007
	The National Post	National	English	
	The Toronto Star	National	English	
	The Calgary Herald	Regional	English	
	The Edmonton Journal	Regional	English	
	The Chronicle Herald	Regional	English	
	The Montreal Gazette	Regional	English	
	The Charlottetown Guardian	Regional	English	
	Halifax Daily News	Regional	English	
	The Ottawa Citizen	Regional	English	
	The Star Phoenix	Regional	English	
	The Telegram	Regional	English	
	The Times Colonist	Regional	English	
	The Vancouver Sun	Regional	English	
	The Winnipeg Sun	Regional	English	
	Yukon News	Regional	English	
Rowe [37]	The Globe and Mail	National	English	2007–2008
	The Toronto Star	National	English	
	The Calgary Herald	Regional	English	
	The Toronto Sun	Regional	English	
	The Montreal Gazette	Regional	English	
	The Ottawa Citizen	Regional	English	
Young and Dugas [38]	The Globe and Mail	National	English	1988–2008
	The National Post	National	English	
Anne DiFrancesco and Young [39]	The Globe and Mail	National	English	2008
	The National Post	National	English	
Ahchong and Dodds [40]	The Globe and Mail	National	English	1988–2007

Continued on next page

Table A1 – *Continued from previous page*

Study	Newspaper(s)	Scope	Language	Time Period
	The Toronto Star	National	English	
Stoddart and Tindall [41]	The Globe and Mail	National	English	1999–2010
	The National Post	National	English	
Ford and King [12]	The Globe and Mail	National	English	1993–2013
	The Toronto Star	National	English	
Stoddart and Smith [42]	The Globe and Mail	National	English	2006–2010
	The National Post	National	English	
Stoddart et al. [43]	The Globe and Mail	National	English	1997–2010
	The National Post	National	English	
Barkemeyer et al. [44]	The Globe and Mail	National	English	2008
	The National Post	National	English	
	The Toronto Star	National	English	
	The Vancouver Sun	Regional	English	
	The Toronto Sun	Regional	English	
King et al. [29]	The Globe and Mail	National	English	2005–2015
	The National Post	National	English	
	The Toronto Star	National	English	
	The Calgary Herald	Regional	English	
	The Chronicle Herald	Regional	English	
	The Vancouver Sun	Regional	English	
	The Whitehorse Daily Star	Regional	English	
	Journal de Montréal	Regional	French	
Pillod [45]	The Globe and Mail	National	English	2008–2020
[46]	The Toronto Star	National	English	2015–2018
	The Calgary Herald	Regional	English	
	The Vancouver Sun	Regional	English	
	The Chronicle Herald	Regional	English	
	The Winnipeg Free Press	Regional	English	
	Times and Transcript	Regional	English	
Hase et al. [47]	The Globe and Mail	National	English	2006–2018

Continued on next page

Table A1 – *Continued from previous page*

Study	**Newspaper(s)**	**Scope**	**Language**	**Time Period**
	The Toronto Star	National	English	
Sachdeva and McCaffrey [48]	n/a	n/a	English	1986–2016
Stoddart and Yang [49]	The Globe and Mail	National	English	2015–2020
	The National Post	National	English	
	The Telegraph	Regional	English	
	The Telegram	Regional	English	
	The Chronicle Herald	Regional	English	
	The Guardian	Regional	English	
Chen et al. [50]	n/a	n/a	English	2018–2021

Note: This table summarizes the newspaper sources, geographic scope, language, and time periods covered by previous studies on climate change media coverage in Canada. Most studies focus exclusively on English-language national newspapers, particularly The Globe and Mail and The National Post.

Table A2: Overview of newspapers included in the CCF database

Province/Territory	**Newspaper**	**Language**	**Avg. Weekday Circ. (2015)**	**Articles**
National	The Globe and Mail	English	226,171	29,170
	The National Post	English	79,157	19,924
	The Toronto Star	English	192,084	46,557
Alberta	The Edmonton Journal	English	59,244	18,102
	The Calgary Herald	English	58,865	19,274
British Columbia	The Vancouver Sun	English	78,901	17,782
	The Times Colonist	English	47,556	11,744
Manitoba	The Winnipeg Free Press	English	67,503	12,296
New Brunswick	L'Acadie Nouvelle	French	18,102	5,067
Newfoundland & Labrador	The Telegram	English	15,380	5,838
Nova Scotia	The Chronicle Herald	English	71,304	10,723
Ontario	The Toronto Sun	English	85,775	3,121
	Le Droit	French	28,616	4,727
Quebec	Le Devoir	French	28,729	13,470
	Le Journal de Montréal	French	175,220	5,381
	La Presse	French	98,657	6,917
	La Presse+[a]	French	–	9,299
	The Montreal Gazette	English	53,344	9,510
Saskatchewan	The Star Phoenix	English	31,081	7,752
Yukon	The Whitehorse Daily Star	English	945	7,603
Total				**264,257**

[a] In 2013, La Presse launched a free digital tablet edition called La Presse+ which offers slightly different content. The corpus considers articles published both on the website and in the digital tablet edition. Duplicates were discarded.

B Appendix B: Complete Framework and Performance Metrics

Table B1: Complete CCF annotation framework with operational definitions and database codes (67 categories*)

#	Category	Code	Description
THEMATIC FRAMES			
Economic Frame			
1	**Economic Frame (Primary Category)**	`economic_frame`	If the sentence refers to climate change as an economic issue, including its negative or positive impacts on the economy, the financial costs or benefits of climate action, or the carbon footprint of the economic or industrial sector. This label captures any framing of climate change in economic terms, such as effects on jobs, markets, industries, or public and private finances.
2	Negative impacts of climate change on the economy	`eco_neg_impact`	If the sentence refers to the negative effects of climate change on the economy, such as damage to infrastructure, loss of agricultural productivity, disruptions to trade, increased costs for businesses, or economic decline in sectors vulnerable to climate shifts. This includes any economic losses directly resulting from changing climate conditions.
3	Positive impacts of climate change on the economy	`eco_pos_impact`	If the sentence refers to the positive effects of climate change on the economy, such as new agricultural opportunities, the opening of new trade routes, or the potential for economic growth in sectors adapting to climate shifts. This includes any economic gains directly resulting from changing climate conditions.
4	Economic disadvantages of climate action	`eco_cost`	If the sentence refers to the negative economic impacts caused by climate policies or climate action, such as costs associated with transitioning to green energy, investments in climate adaptation, debt incurred from financing climate action, or short-term job losses in fossil fuel industries. This category captures financial burdens and challenges linked to climate action or policy.
5	Economic benefits of climate action	`eco_benefit`	If the sentence refers to the positive economic impacts resulting from climate policies or climate action, such as job creation in renewable energy sectors, investments in green technologies, economic growth driven by sustainability initiatives, or cost savings from energy efficiency. This category captures financial gains and opportunities linked to climate action or policy.

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
6	Carbon footprint of the economic sector	`eco_footprint`	If the sentence refers to the environmental or carbon footprint of the economic or industrial sectors, including emissions and resource use related to manufacturing, transportation, energy production, mining, agriculture, and other commercial or industrial activities. This label covers the ecological impact generated by economic and industrial processes, not the effects of climate change on these sectors.
Health Frame			
7	**Health Frame (Primary Category)**	`health_frame`	If the sentence refers to climate change as a public health issue, including its negative or positive impacts on physical or mental health, the health co-benefits of climate action (e.g., cleaner air, active transportation), or the carbon footprint of the health sector. This label captures any framing of climate change in relation to health outcomes, healthcare systems, or health-related consequences and solutions.
8	Negative impacts of climate change on health	`health_neg_impact`	If the sentence refers to the negative effects of climate change on human health, such as increased risks of infectious diseases (e.g., malaria, dengue), respiratory and cardiovascular illnesses due to air pollution and heat exposure, heat-related illnesses (e.g., heatstroke), and mental health disorders (e.g., anxiety, depression) linked to environmental changes.
9	*Positive impacts of climate change on health**	`health_pos_impact`	If the sentence refers to the potential positive effects of climate change on human health, such as reductions in cold-related illnesses and deaths, or increased opportunities for outdoor activity due to milder winters.
10	Health co-benefits of climate action	`health_cobenefit`	If the sentence refers to health benefits resulting from climate policies or actions, such as improved air quality, better nutrition, reduced respiratory or cardiovascular diseases, fewer premature deaths, and other positive health outcomes linked to climate mitigation and adaptation efforts.
11	*Carbon footprint of the health sector**	`health_footprint`	If the sentence refers to the environmental or carbon footprint of the health sector, including emissions and resource use associated with healthcare facilities, medical equipment, pharmaceuticals production, patient and staff travel, and waste generated by medical activities. This label covers the ecological impact of delivering health services, not the impacts of climate change on health outcomes.
Security Frame			
12	**Security Frame (Primary Category)**	`security_frame`	If the sentence refers to the effects of climate change on security, such as violent conflict driven by resource scarcity, an influx of climate refugees, disruption of military operations, or the carbon footprint of the defense sector. This category includes broader security challenges arising from climate change, including geopolitical tensions or the strategic responses to climate impacts.

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
13	Presence of climate refugees	`security_refugees`	If the sentence refers to the displacement of populations or the presence of climate refugees as a result of a natural disaster, environmental degradation, or climate change impact. This category captures the movement of people fleeing climate-related disasters.
14	Conflict	`security_conflict`	If the sentence refers to the emergence or intensification of conflicts over natural resource exploitation due to climate change, such as disputes over water, land, or energy resources exacerbated by shifting climate patterns. This includes conflicts arising from scarcity or competition for increasingly stressed resources.
15	*Post-disaster military assistance**	`security_military`	If the sentence refers to the deployment of military forces as reinforcement after a natural disaster, such as wildfires, floods, or other extreme events. This includes military support in disaster response and recovery, like providing humanitarian aid or maintaining order during a crisis.
16	*Disruption of military operations**	`security_disruption`	If the sentence refers to the disruption of military operations due to a natural disaster affecting a military base or facilities. This includes cases where climate-induced events damage or disrupt military infrastructure, affecting defense readiness or response capabilities.
Justice Frame			
17	**Justice Frame (Primary Category)**	`justice_frame`	If the sentence frames climate change as a justice or moral issue, including references to who benefits or loses from climate action, differentiated responsibilities across countries or groups, unequal impacts of climate change, disparities in access to mitigation or adaptation measures, or concerns about intergenerational justice. This label captures ethical, equity-based, or fairness-oriented dimensions of climate discourse.
18	Winners and losers of climate action	`justice_winners`	If the sentence refers to how climate policies or actions create winners and losers, by benefiting certain groups (e.g., vulnerable populations, Indigenous communities, developing countries) while disadvantaging others (e.g., fossil fuel workers, carbon-intensive industries, or developed nations facing steeper transitions). This label focuses on distributional outcomes of climate policy.
19	Differentiated responsibility	`justice_responsibility`	If the sentence refers to the unequal responsibility for causing climate change, especially in reference to the principle of common but differentiated responsibilities. This includes statements emphasizing that developed countries, multinational corporations, or high-emitting individuals have historically contributed more to global warming and should therefore bear a greater share of the mitigation burden.

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
20	Unequal vulnerability to climate change	`justice_vulnerability`	If the sentence highlights how different populations are unequally affected by the impacts of climate change, due to geographic, social, or economic vulnerabilities. This includes references to marginalized groups such as women, seniors, racialized populations, people in developing countries, or those living in high-risk zones (e.g., coastal or arid regions). This label focuses on impacts, not responsibility or capacity.
21	Unequal access to climate action	`justice_access`	If the sentence refers to unequal capacity to act on climate change, emphasizing that countries, companies, or individuals do not have equal financial, technical, or institutional means to take climate action. For example, small businesses may struggle to decarbonize compared to multinational firms, and low-income households may not afford sustainable alternatives. This label reflects inequality in ability to act, not impacts or blame.
22	Intergenerational justice	`justice_intergen`	If the sentence refers to intergenerational justice, meaning that current decisions and actions should not compromise the rights or well-being of future generations. Do not apply this label to sentences that merely mention age differences or generational groups (e.g., "young people are protesting") unless the statement clearly relates to moral responsibility across generations or the long-term consequences of climate inaction.
Political Frame			
23	**Political Frame (Primary Category)**	`political_frame`	If the sentence frames climate change as a political issue, including references to policy adoption or implementation, political debates or controversies, political positioning by parties or leaders on climate issues, and public opinion data related to climate change. This label captures the role of governance, decision-making, and political dynamics in climate discourse.
24	Policy action	`pol_action`	If the sentence refers to the formal approval, adoption, or enactment of a policy, plan, or strategy by a government at the local, provincial, national, or international level. This includes the passing of new legislation, regulations, government programs, or official commitments related to climate or environmental issues (e.g., approval of a carbon pricing law, ratification of a renewable energy policy, adoption of a climate action framework).
25	Political debate	`pol_debate`	If the sentence refers to a political debate or discussion about climate change, particularly focusing on the differing views or disagreements on climate policies, the effectiveness of past actions, or the responsibility of different levels of government (e.g., federal vs. provincial). This includes disagreements about specific climate initiatives, such as proposed or enacted policies, laws, or funding allocations.

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
26	Political positioning	`pol_position`	If the sentence refers to the explicit stance or position of a politician or political party on climate change or climate-related policies. This includes statements where a political figure or party expresses their support or opposition to specific climate actions (e.g., carbon taxes, energy transitions, or climate mitigation strategies), or when they make electoral promises regarding climate policies.
27	Public opinion data	`pol_opinion`	If the sentence refers to data, surveys, polls, or studies measuring public opinion, attitudes, beliefs, or perceptions related to climate change, environmental policies, or climate action. This label captures quantitative or qualitative information on how the public views climate issues.
Scientific Frame			
28	**Scientific Frame (Primary Category)**	`scientific_frame`	If the sentence frames climate change as a scientific issue, including references to scientific research, discoveries, explanations, or assessments of climate change. This includes mentions of scientific consensus or controversy, as well as expressions of uncertainty or certainty in climate science. This label captures the role of scientific knowledge and expertise in understanding and communicating climate change.
29	Scientific debate	`sci_debate`	If the sentence refers to a past or ongoing debate or disagreement within the scientific community about aspects of climate science or climate-related technologies. This includes contested interpretations of data, disputes about methodologies, or ethical debates surrounding proposed scientific or technological solutions such as geoengineering. It does not include general questioning of climate science or statements that challenge the validity or sufficiency of climate evidence.
30	Popularisation or scientific discovery	`sci_discovery`	If the sentence explains or describes climate science mostly grounded in the natural sciences—such as chemistry, physics, or biology—or presents scientific or technological discoveries related to climate change. This includes explanations of mechanisms (e.g., the greenhouse effect, climate feedbacks), climate modeling, or innovations such as carbon capture. It does not apply to statements that merely describe environmental conditions or trends without explaining an underlying scientific process.
31	Questioning of climate science	`sci_skepticism`	If the sentence questions the validity, accuracy, or sufficiency of climate science. This includes claims that climate projections are flawed, evidence is lacking, or conclusions are exaggerated. It only applies when science is being challenged or delegitimized, not when scientists acknowledge limitations, complexity, or areas of ongoing research.

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
32	Defense of climate science	`sci_defense`	If the sentence defends the robustness or credibility of climate science in response to skepticism or denial. This includes assertions that reinforce the scientific consensus, affirm the validity of climate models, or rebut claims that climate science is incorrect or misleading. It only applies when climate science is defended, not when it is neutrally presented.
Environmental Frame			
33	**Environmental Frame (Primary Category)**	`environmental_frame`	If the sentence frames climate change as an environmental issue, including references to the degradation or disappearance of natural habitats, loss of biodiversity, and negative impacts on fauna and flora. This label captures ecological consequences of climate change and emphasizes the relationship between climate disruption and the natural environment.
34	Loss of natural environments	`env_habitat`	If the sentence refers to the long-term or irreversible disappearance or degradation of natural environments due to climate change. This includes situations where ecosystems or environmental features (e.g., glaciers, coral reefs) are predicted or expected to permanently disappear or significantly decline as a result of climate impacts, indicating that they will not recover. It does not apply to short-term, isolated, or recoverable events, such as one-time disasters (e.g., a single wildfire or a temporary drought) that might affect natural environments but do not necessarily imply permanent loss.
35	Loss of fauna and flora	`env_species`	If the sentence refers to the negative impacts of climate change on animal and plant species, including changes in habitat, shifts in migration or breeding patterns, increased mortality, loss of biodiversity, or threats to species survival. This label covers both temporary and ongoing effects on fauna and flora caused by climate change.
Cultural Frame			
36	**Cultural Frame (Primary Category)**	`cultural_frame`	If the sentence frames climate change as a cultural issue, including cultural depictions of climate change through art, literature, music, or film; challenges to hosting artistic or sports events due to climate impacts; the loss or disruption of Indigenous cultural practices; or the carbon footprint of the cultural and creative sectors. This label captures how climate change affects and is expressed through cultural life and identities.
37	Artistic representation	`cult_art`	If the sentence refers to cultural representations of climate change in various forms such as art, literature, music, film, documentaries, theater, or other creative expressions. This label captures how climate change is portrayed, interpreted, or communicated through cultural and artistic mediums.

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
38	Difficulty to host cultural or sports events	`cult_event_impact`	If the sentence refers to how climate change disrupts or challenges the organization of artistic, cultural, or sports events, such as cancelled festivals, concerts, or competitions due to extreme weather (e.g., storms, heatwaves, wildfires, or lack of snow for winter sports). This label captures climate-related obstacles to event planning or continuity.
39	Loss of indigenous practices	`cult_indigenous`	If the sentence refers to the erosion or disruption of traditional cultural practices among Indigenous communities due to climate change, such as changes in animal migration, plant cycles, or land access that affect food systems, ceremonies, or knowledge transmission. This label emphasizes climate-driven impacts on Indigenous cultural continuity.
40	Carbon footprint of the cultural and sports sectors	`cult_footprint`	If the sentence refers to the environmental or carbon footprint of the cultural, artistic, or sports sectors, including the emissions and resource use associated with activities such as touring, international travel, energy-intensive venues, unsustainable stage or set production, or event-related waste. This label covers the ecological impact of producing and hosting cultural or sports events, not the impacts of climate on those events.
PRIMARY CATEGORIES			
Actors/Messengers			
41	**Presence of Messengers (Primary Category)**	`messenger`	If the sentence includes mentions or quotes of persons or organizations by journalists, whether through direct quotations or indirect references. This label captures the presence of sources, experts, officials, activists, or other actors cited or named within the text. A single individual or organization may be associated with multiple types of expertise or roles (e.g., a doctor would fall both under Health expert and Natural scientist).
42	Health expert	`msg_health`	If the quoted or reported statement comes from a person or organization with medical or health-related recognized expertise. This includes individuals working in the fields of medicine, public health, healthcare administration, or biomedical research, whether in a professional, academic, or governmental capacity (e.g., professor of public health, doctor, nurse, family physician, public health officer, WHO representative, Red Cross medical coordinator, Doctors without Borders spokesperson, Canadian Medical Association representative, certified health consultants, minister of Health).

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
43	Economic expert	`msg_economic`	If the quoted or reported statement comes from a person or organization with economic or financial recognized expertise. This includes individuals working in the fields of economics, finance, business, or economic policy, whether in a professional, academic, private sector, or governmental capacity (e.g., economist, finance minister, central bank official, business executive, investment analyst, professor of economics, representative of a chamber of commerce, World Bank or IMF spokesperson, economic think tank expert, labour market analyst).
44	Security expert	`msg_security`	If the quoted or reported statement comes from a person or organization with recognized expertise in security or defence matters. This includes individuals with academic, professional, or operational experience in military, national security, intelligence, or strategic affairs (e.g., professor in defense or security studies, active or retired member of the armed forces, intelligence agency representative, conflict resolution specialist, national defence minister, security analyst, representative of a security-focused think tank or NGO).
45	Legal expert	`msg_legal`	If the quoted or reported statement comes from a person or organization with recognized expertise in law or legal affairs. This includes individuals working in legal professions, legal academics, the judiciary, or justice-related government roles (e.g., lawyer, prosecutor, law professor, minister of justice, attorney general, legal aid representative, bar association representative, legal consultants).
46	Cultural or Sport expert	`msg_cultural`	If the quoted or reported statement comes from a person with recognized expertise in sports, arts, or culture (e.g., Olympian, coach, professor of literature, museum director, filmmaker, writer, festival organizer, musician, artist, indigenous knowledge keeper).
47	Natural scientist	`msg_scientist`	If the quoted or reported statement comes from a person or organization with recognized expertise in the natural or hard sciences. This includes individuals working in fields such as climatology, environmental science, physics, chemistry, biology, earth sciences, engineering, or related disciplines, whether in academic, governmental, or research institutions (e.g., climate scientist, environmental researcher, oceanographer, physicist, ecologist, geologist, engineer, professor of atmospheric sciences, IPCC contributor, Environment Canada scientist, research institute spokesperson).
48	Social scientist	`msg_social`	If the quoted or reported statement comes from a person with expertise in social sciences, helping to understand how society perceives, responds to, and governs issues like climate change (e.g., sociologist, political scientist, media scholar, economist).

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
49	Activist	`msg_activist`	If the quoted or reported statement comes from a person or organization known primarily for advocacy or activism. This includes individuals or groups engaged in campaigning, protesting, or raising awareness on environmental, social, or climate-related issues, whether through grassroots organizing, NGOs, or international movements (e.g., climate activist, environmental campaigner, Indigenous land defender, Fridays for Future organizer, Greenpeace spokesperson, Extinction Rebellion member, citizen activist, youth climate leader, environmental justice advocate).
50	Public official	`msg_official`	If the quoted or reported statement comes from a politician, elected official, government representative, or public decision-maker at any level of government. This includes individuals holding political office or involved in policymaking, regulation, or governance, whether municipal, provincial, federal, or international (e.g., prime minister, environment minister, member of parliament, city councillor, premier, senator, government spokesperson, cabinet member, UN climate negotiator, provincial regulator, parliamentary committee chair).
Events			
51	**Presence of Events (Primary Category)**	`event`	If the sentence refers to the occurrence or mention of one or more events. This includes extreme meteorological events (e.g., storms, floods, heatwaves), political or governmental events (e.g., elections, policy announcements, trials, judiciary decisions), meetings or conferences, publications, cultural events, protests, or other significant happenings.
52	Extreme meteorological event	`evt_weather`	If the sentence refers to the occurrence of an extreme meteorological or climate-related event that causes or risks causing significant disruption, damage, or harm. This includes events such as heatwaves, floods, wildfires, hurricanes, droughts, storms, or other weather-related disasters, whether sudden or prolonged.
53	Meeting	`evt_meeting`	If the sentence refers to an official visit, conference, summit, forum, meeting, convention, or assembly that brings together people to discuss, negotiate, or present on specific topics at the local, national, or international levels (e.g., town hall meeting, premiers' conference, ministerial roundtable, state visit, UN conference, international forum on clean energy, G7 meeting, COP28, World Economic Forum).
54	Publication	`evt_publication`	If the sentence refers to the release or publication of a document such as a report, academic article, survey results, op-ed, policy brief, or white paper, often aimed at informing the public, guiding decision-making, or contributing to public discourse (e.g., IPCC report, Statistics Canada survey, university study on climate impacts, editorial on climate policy, NGO report on emissions, think tank publication).

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
55	Election	`evt_election`	If the sentence refers to the occurrence of an election at the local, provincial, national, or international level, including general elections, by-elections, leadership races, referenda, or electoral campaigns (e.g., federal election, municipal by-election, provincial leadership race, EU parliamentary elections).
56	New policy	`evt_policy`	If the sentence refers to the announcement or unveiling of a policy, plan, or strategy at the local, provincial, national, or international level, including new legislation, regulations, government programs, or official commitments, including those related to climate or environmental issues (e.g., carbon pricing plan, renewable energy strategy, climate action framework).
57	Judiciary decision	`evt_judiciary`	If the sentence refers to the occurrence of a trial, court ruling, legal proceeding, or judicial decision at any level of the judicial system, including constitutional challenges, environmental lawsuits, regulatory hearings, or Supreme Court decisions, including those related to climate or environmental issues (e.g., court ruling on carbon pricing, environmental class action lawsuit, constitutional challenge to climate legislation, tribunal decision on pipeline approval).
58	Cultural or Sports event	`evt_cultural`	If the sentence refers to the organization or hosting of a sports, artistic, or cultural event (e.g., Olympics games, national hockey tournament, local marathon, film screening, music concert, theatre performance, book launch, mural festival, photography exhibit).
59	Protest	`evt_protest`	If the sentence refers to the organization of a protest (e.g., climate strike, anti-pipeline protest, extinction rebellion action, anti-racism march, women's right rally, union-led protest for better wages).
Solutions			
60	**Presence of Solutions (Primary Category)**	`solution`	If the sentence refers to solutions addressing climate change, including mitigation efforts to reduce greenhouse gas emissions or adaptation strategies to manage climate impacts. This includes direct mentions of specific policies, technologies, behaviors, or initiatives, as well as indirect references to approaches aimed at combating or coping with climate change.
61	Mitigation strategy	`sol_mitigation`	If the sentence refers to mitigation solutions designed to reduce or prevent the emission of greenhouse gases and limit the extent of future climate change. These strategies aim to address the root causes of climate change by lowering carbon footprints, enhancing carbon sinks, or transitioning to low-carbon technologies (e.g., investing in renewable energy, improving energy efficiency, implementing carbon pricing, promoting public transit, reforestation, phasing out fossil fuels).

Continued on next page

Table B1 – *Continued from previous page*

#	Category	Code	Description
62	Adaptation strategy	`sol_adaptation`	If the sentence refers to adaptation solutions designed to reduce vulnerability to the negative effects of climate change by adjusting or preparing for those impacts. These strategies involve actions to manage the risks and enhance resilience, particularly by reducing exposure to climate-related hazards and minimizing harm to people, infrastructure, and ecosystems (e.g., providing cooling centres during heat waves, restoring wetlands to limit flooding, altering farming techniques to cope with changing weather patterns).
EMOTIONAL TONE			
63	Emotion: positive	`tone_positive`	If the tone is generally optimistic, reassuring, or enthusiastic, highlighting favorable aspects or expressing confidence in climate solutions or actions.
64	Emotion: negative	`tone_negative`	If the tone is generally alarming, critical, or pessimistic, especially if the sentence expresses concern, fear, or frustration about climate change or its impacts.
65	Emotion: neutral	`tone_neutral`	If the sentence is mainly informative or descriptive without an evaluative tone, presenting facts or balanced perspectives without strong emotional content.
GEOGRAPHIC FOCUS			
66	Mention of Canada	`canada`	If the sentence refers to Canada in any context, domestic or international, including but not limited to climate change impacts on Canadian territory, as well as Canada's policies, actions, or roles.
URGENCY/ALARMISM			
67	Urgency to act	`urgency`	If the sentence conveys a strong sense of urgency or alarmism, emphasizing the immediate need for action or highlighting critical risks and dangers related to climate change, often with a tone that signals warning or crisis.

* Categories marked with an asterisk were included in the annotation protocol but had insufficient positive examples for model training. Health subcategories (`health_pos_impact`, `health_footprint`) were completely excluded due to insufficient annotations in both English and French. Security subcategories (`security_military`, `security_disruption`) are partially available (trained in some languages only). The framework includes 67 categories in total: 63 fully available, 2 partially available, and 2 excluded.

Table B2: Complete model training performance metrics for all annotation categories

#	Category	Code	F1 (Class 1)		F1 (Class 0)		Macro F1	
			EN	FR	EN	FR	EN	FR
THEMATIC FRAMES								
Economic Frame								
1	**Economic Frame (Primary Category)**	`economic_frame`	0.745	0.814	0.944	0.957	0.845	0.885
2	Negative impacts of climate change on the economy	`eco_neg_impact`	0.727	0.889	0.914	0.944	0.821	0.917
3	Positive impacts of climate change on the economy	`eco_pos_impact`	0.333	0.400	0.995	0.996	0.664	0.698
4	Economic disadvantages of climate action	`eco_cost`	0.615	0.500	0.848	0.913	0.732	0.707
5	Economic benefits of climate action	`eco_benefit`	0.500	0.516	0.952	0.981	0.726	0.749
6	Carbon footprint of the economic sector	`eco_footprint`	0.857	0.857	0.938	0.950	0.897	0.904
Health Frame								
7	**Health Frame (Primary Category)**	`health_frame`	0.800	0.667	0.989	0.994	0.894	0.830
8	Negative impacts of climate change on health	`health_neg_impact`	0.909	0.857	0.000	0.000	0.455	0.429
9	Health co-benefits of climate action	`health_cobenefit`	0.400	0.011	0.571	0.227	0.486	0.119
Security Frame								
10	**Security Frame (Primary Category)**	`security_frame`	0.870	0.800	0.996	0.997	0.933	0.898

Continued on next page

Table B2 – *Continued from previous page*

#	Category	Code	F1 (Class 1)		F1 (Class 0)		Macro F1	
			EN	**FR**	**EN**	**FR**	**EN**	**FR**
11	Presence of climate refugees	`security_refugees`	1.000	1.000	1.000	1.000	1.000	1.000
12	Conflict	`security_conflict`	1.000	1.000	1.000	1.000	1.000	1.000
13	Post-disaster military assistance	`security_military`	–	0.010	–	0.043	–	0.026
14	Disruption of military operations	`security_disruption`	–	1.000	–	1.000	–	1.000
Justice Frame								
15	**Justice Frame (Primary Category)**	`justice_frame`	0.719	0.717	0.975	0.981	0.847	0.849
16	Winners and losers of climate action	`justice_winners`	0.667	0.667	0.933	0.923	0.800	0.795
17	Differentiated responsibility	`justice_responsibility`	0.857	0.750	0.909	0.750	0.883	0.750
18	Unequal vulnerability to climate change	`justice_vulnerability`	0.571	0.600	0.992	0.995	0.782	0.798
19	Unequal access to climate action	`justice_access`	0.364	0.625	0.000	0.993	0.182	0.809
20	Intergenerational justice	`justice_intergen`	0.933	0.800	0.999	0.999	0.966	0.899
Political Frame								
21	**Political Frame (Primary Category)**	`political_frame`	0.808	0.774	0.897	0.888	0.853	0.831
22	Policy action	`pol_action`	0.621	0.667	0.985	0.966	0.803	0.816
23	Political debate	`pol_debate`	0.916	0.966	0.222	0.571	0.569	0.768
24	Political positioning	`pol_position`	0.750	0.800	0.946	0.989	0.848	0.895
25	Public opinion data	`pol_opinion`	0.909	1.000	0.999	1.000	0.954	1.000
Scientific Frame								
26	**Scientific Frame (Primary Category)**	`scientific_frame`	0.784	0.702	0.953	0.962	0.869	0.832

Continued on next page

Table B2 – *Continued from previous page*

#	Category	Code	F1 (Class 1)		F1 (Class 0)		Macro F1	
			EN	**FR**	**EN**	**FR**	**EN**	**FR**
27	Scientific debate	`sci_debate`	0.737	0.500	0.848	0.875	0.793	0.688
28	Popularisation or scientific discovery	`sci_discovery`	0.828	0.938	0.800	0.750	0.814	0.844
29	Questioning of climate science	`sci_skepticism`	0.727	0.333	0.927	0.995	0.827	0.664
30	Defense of climate science	`sci_defense`	0.571	0.000	0.933	0.919	0.752	0.459
Environmental Frame								
31	**Environmental Frame (Primary Category)**	`environmental_frame`	0.842	0.625	0.989	0.980	0.915	0.802
32	Loss of natural environments	`env_habitat`	1.000	0.857	1.000	0.800	1.000	0.829
33	Loss of fauna and flora	`env_species`	0.889	0.857	0.857	0.800	0.873	0.829
Cultural Frame								
34	**Cultural Frame (Primary Category)**	`cultural_frame`	0.773	0.833	0.986	0.993	0.879	0.913
35	Artistic representation	`cult_art`	1.000	1.000	1.000	1.000	1.000	1.000
36	Difficulty to host cultural or sports events	`cult_event_impact`	0.706	1.000	0.993	1.000	0.850	1.000
37	Loss of indigenous practices	`cult_indigenous`	1.000	0.026	1.000	0.899	1.000	0.462
38	Carbon footprint of the cultural and sports sectors	`cult_footprint`	1.000	0.005	1.000	0.143	1.000	0.074
PRIMARY CATEGORIES								
Actors/Messengers								
39	**Presence of Messengers (Primary Category)**	`messenger`	0.912	0.929	0.904	0.915	0.908	0.922
40	Health expert	`msg_health`	0.857	0.909	0.997	0.999	0.927	0.954

Continued on next page

Table B2 – *Continued from previous page*

#	Category	Code	F1 (Class 1)		F1 (Class 0)		Macro F1	
			EN	**FR**	**EN**	**FR**	**EN**	**FR**
41	Economic expert	`msg_economic`	0.600	0.750	0.970	0.973	0.785	0.861
42	Security expert	`msg_security`	0.667	1.000	0.993	1.000	0.830	1.000
43	Legal expert	`msg_legal`	0.667	0.769	0.999	0.996	0.833	0.883
44	Cultural or Sport expert	`msg_cultural`	1.000	0.556	1.000	0.990	1.000	0.773
45	Natural scientist	`msg_scientist`	0.789	0.833	0.977	0.971	0.883	0.902
46	Social scientist	`msg_social`	0.769	0.645	0.996	0.986	0.883	0.816
47	Activist	`msg_activist`	0.571	1.000	0.978	1.000	0.775	1.000
48	Public official	`msg_official`	0.703	0.923	0.897	0.964	0.800	0.943
Events								
49	**Presence of Events (Primary Category)**	`event`	0.794	0.819	0.935	0.932	0.865	0.876
50	Extreme meteorological event	`evt_weather`	1.000	0.857	1.000	0.968	1.000	0.912
51	Meeting	`evt_meeting`	0.824	0.957	0.936	0.982	0.880	0.969
52	Publication	`evt_publication`	0.873	1.000	0.990	1.000	0.931	1.000
53	Election	`evt_election`	1.000	0.800	1.000	0.998	1.000	0.899
54	New policy	`evt_policy`	0.769	0.727	0.943	0.954	0.856	0.841
55	Judiciary decision	`evt_judiciary`	1.000	1.000	1.000	1.000	1.000	1.000
56	Cultural or Sports event	`evt_cultural`	0.333	1.000	0.995	1.000	0.664	1.000
57	Protest	`evt_protest`	0.889	0.727	0.999	0.996	0.944	0.862
Solutions								
58	**Presence of Solutions (Primary Category)**	`solution`	0.737	0.878	0.914	0.944	0.825	0.911
59	Mitigation strategy	`sol_mitigation`	0.750	0.812	0.935	0.942	0.842	0.877
60	Adaptation strategy	`sol_adaptation`	0.696	0.800	0.991	0.989	0.843	0.894
EMOTIONAL TONE								
61	Emotion: positive	`tone_positive`	0.526	0.690	0.966	0.968	0.746	0.829
62	Emotion: negative	`tone_negative`	0.706	0.765	0.785	0.843	0.745	0.804
63	Emotion: neutral	`tone_neutral`	0.741	0.789	0.676	0.693	0.709	0.741

Continued on next page

Table B2 – *Continued from previous page*

#	Category	Code	F1 (Class 1)		F1 (Class 0)		Macro F1	
			EN	**FR**	**EN**	**FR**	**EN**	**FR**
GEOGRAPHIC FOCUS								
64	Mention of Canada	`canada`	0.942	0.964	0.968	0.980	0.955	0.972
URGENCY TO ACT								
65	Urgency to act	`urgency`	0.591	0.649	0.975	0.984	0.783	0.816
AVERAGE PERFORMANCE METRICS								
		English Average	0.769	–	0.905	–	0.837	–
		French Average	–	0.737	–	0.894	–	0.816
		Overall Average	0.753		0.900		0.826	
TOTAL: 65 ANNOTATION CATEGORIES *(65 categories with at least one model; 0 categories entirely excluded*)*								

* Insufficient training data (fewer than 10 positive examples in training set).

Table B3: Training and validation dataset distribution across all annotation categories

#	Category	Code	English				French			
			Training		Validation		Training		Validation	
			Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
THEMATIC FRAMES										
Economic Frame										
1	**Economic Frame (Primary Category)**	`economic_frame`	218	1067	24	118	252	1164	28	129
2	Negative impacts of climate change on the economy	`eco_neg_impact`	61	158	6	17	75	178	8	19
3	Positive impacts of climate change on the economy	`eco_pos_impact`	14	205	1	22	7	245	1	27
4	Economic disadvantages of climate action	`eco_cost`	63	156	6	17	42	211	4	23
5	Economic benefits of climate action	`eco_benefit`	36	183	3	20	48	205	5	22
6	Carbon footprint of the economic sector	`eco_footprint`	67	152	7	16	77	176	8	19
Health Frame										
7	**Health Frame (Primary Category)**	`health_frame`	57	1228	6	136	37	1379	4	153
8	Negative impacts of climate change on health	`health_neg_impact`	47	10	5	1	32	5	3	1
9	Health co-benefits of climate action	`health_cobenefit`	8	49	1	5	4	33	1	3
Security Frame										
10	**Security Frame (Primary Category)**	`security_frame`	19	1266	2	140	19	1397	2	155

Continued on next page

Table B3 – *Continued from previous page*

#	Category	Code	English				French			
			Training		Validation		Training		Validation	
			Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
11	Presence of climate refugees	`security_refugees`	10	9	1	1	7	12	1	1
12	Conflict	`security_conflict`	6	13	1	1	7	12	1	1
13	Post-disaster military assistance	`security_military`	0	18	1	2	4	15	1	1
14	Disruption of military operations	`security_disruption`	0	18	1	2	1	18	1	1
Justice Frame										
15	**Justice Frame (Primary Category)**	`justice_frame`	90	1196	9	132	81	1335	9	148
16	Winners and losers of climate action	`justice_winners`	20	70	2	7	20	62	2	6
17	Differentiated responsibility	`justice_responsibility`	38	52	4	5	41	41	4	4
18	Unequal vulnerability to climate change	`justice_vulnerability`	8	81	1	9	9	72	1	8
19	Unequal access to climate action	`justice_access`	21	69	2	7	20	62	2	6
20	Intergenerational justice	`justice_intergen`	14	76	1	8	6	75	1	8
Political Frame										
21	**Political Frame (Primary Category)**	`political_frame`	416	869	46	96	440	977	48	108
22	Policy action	`pol_action`	62	355	6	39	58	382	6	42
23	Political debate	`pol_debate`	344	72	38	8	396	45	43	4
24	Political positioning	`pol_position`	61	356	6	39	35	405	3	45
25	Public opinion data	`pol_opinion`	21	396	2	43	20	420	2	46
Scientific Frame										

Continued on next page

Table B3 – *Continued from previous page*

#	Category	Code	English				French			
			Training		Validation		Training		Validation	
			Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
26	**Scientific Frame (Primary Category)**	`scientific_frame`	243	1042	27	115	185	1232	20	136
27	Scientific debate	`sci_debate`	98	146	10	16	46	139	5	15
28	Popularisation or scientific discovery	`sci_discovery`	126	117	14	13	144	41	16	4
29	Questioning of climate science	`sci_skepticism`	50	194	5	21	17	169	1	18
30	Defense of climate science	`sci_defense`	37	207	4	22	32	153	3	17
Environmental Frame										
31	**Environmental Frame (Primary Category)**	`environmental_frame`	84	1201	9	133	63	1354	6	150
32	Loss of natural environments	`env_habitat`	50	35	5	3	35	28	3	3
33	Loss of fauna and flora	`env_species`	42	43	4	4	32	31	3	3
Cultural Frame										
34	**Cultural Frame (Primary Category)**	`cultural_frame`	44	1242	4	137	45	1371	5	152
35	Artistic representation	`cult_art`	20	24	2	2	18	27	2	3
36	Difficulty to host cultural or sports events	`cult_event_impact`	11	33	1	3	21	25	2	2
37	Loss of indigenous practices	`cult_indigenous`	10	34	1	3	4	41	1	4
38	Carbon footprint of the cultural and sports sectors	`cult_footprint`	1	42	1	4	2	43	1	4
PRIMARY CATEGORIES										
Actors/Messengers										

Continued on next page

Table B3 – *Continued from previous page*

#	Category	Code	English				French			
			Training		Validation		Training		Validation	
			Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
39	**Presence of Messengers (Primary Category)**	`messenger`	657	629	72	69	736	681	81	75
40	Health expert	`msg_health`	10	647	1	71	8	728	1	80
41	Economic expert	`msg_economic`	55	602	6	66	67	669	7	74
42	Security expert	`msg_security`	3	653	1	72	10	726	1	80
43	Legal expert	`msg_legal`	1	655	1	72	4	731	1	81
44	Cultural or Sport expert	`msg_cultural`	13	644	1	71	9	727	1	80
45	Natural scientist	`msg_scientist`	107	550	11	61	117	620	12	68
46	Social scientist	`msg_social`	9	648	1	71	26	711	2	78
47	Activist	`msg_activist`	53	604	5	67	55	681	6	75
48	Public official	`msg_official`	171	486	18	54	223	513	24	57
Events										
49	**Presence of Events (Primary Category)**	`event`	298	987	33	109	355	1062	39	117
50	Extreme meteorological event	`evt_weather`	72	226	8	25	59	297	6	32
51	Meeting	`evt_meeting`	75	224	8	24	100	255	11	28
52	Publication	`evt_publication`	74	225	8	24	114	242	12	26
53	Election	`evt_election`	16	283	1	31	18	337	2	37
54	New policy	`evt_policy`	55	243	6	27	60	296	6	32
55	Judiciary decision	`evt_judiciary`	10	288	1	32	10	345	1	38
56	Cultural or Sports event	`evt_cultural`	2	296	1	32	11	344	1	38
57	Protest	`evt_protest`	11	288	1	31	8	347	1	38
Solutions										
58	**Presence of Solutions (Primary Category)**	`solution`	314	972	34	107	415	1001	46	111
59	Mitigation strategy	`sol_mitigation`	279	35	31	3	360	55	40	6

Continued on next page

Table B3 – *Continued from previous page*

#	Category	Code	English				French			
			Training		Validation		Training		Validation	
			Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
60	Adaptation strategy	`sol_adaptation`	46	268	5	29	64	351	7	39
EMOTIONAL TONE										
61	Emotion: positive	`tone_positive`	103	1182	11	131	151	1266	16	140
62	Emotion: negative	`tone_negative`	491	794	54	88	537	880	59	97
63	Emotion: neutral	`tone_neutral`	686	599	76	66	726	691	80	76
GEOGRAPHIC FOCUS										
64	Mention of Canada	`canada`	445	840	49	93	493	924	54	102
URGENCY TO ACT										
65	Urgency to act	`urgency`	66	1219	7	135	63	1354	6	150

Note: Categories with insufficient positive training samples (marked with * in Table B2) were excluded from final model training and are not shown in this distribution table.

Table B4: Named Entity Recognition (`ner_entities` variable) model performance metrics

Language	**Model**	**PER F1**	**ORG F1**	**LOC F1**
English	BERT-base-NER	0.961	0.811	0.925
French	spaCy fr_core_news_lg	0.880	–	–
	CamemBERT-NER	–	0.824	0.929

Note: The hybrid approach for French combines spaCy for person entities (PER) with CamemBERT-NER for organization (ORG) and location (LOC) entities based on empirical evaluation on the dataset. Performance metrics are from the original model documentation.

Table B5: Detailed validation performance metrics for trained models (65 categories)

#	Category	Code	F1 Macro			F1 Micro			F1 Weighted			Support		
			EN	FR	ALL	EN	FR	ALL	EN	FR	ALL	EN	FR	ALL
THEMATIC FRAMES														
Economic Frame														
1	**Economic Frame (Primary Category)**	`economic_frame`	0.847	0.814	0.830	0.892	0.864	0.878	0.890	0.862	0.875	120	129	249
2	Negative impacts of climate change on the economy	`eco_neg_impact`	0.820	0.741	0.783	0.848	0.810	0.828	0.856	0.833	0.842	26	18	44
3	Positive impacts of climate change on the economy	`eco_pos_impact`	0.806	0.508	0.684	0.905	0.845	0.873	0.913	0.908	0.900	12	1	13
4	Economic disadvantages of climate action	`eco_cost`	0.741	0.702	0.723	0.800	0.802	0.801	0.813	0.806	0.810	22	23	45
5	Economic benefits of climate action	`eco_benefit`	0.885	0.817	0.848	0.924	0.871	0.896	0.928	0.872	0.899	19	26	45
6	Carbon footprint of the economic sector	`eco_footprint`	0.785	0.788	0.787	0.838	0.845	0.842	0.831	0.843	0.837	30	29	59
Health Frame														
7	**Health Frame (Primary Category)**	`health_frame`	0.800	0.817	0.808	0.947	0.961	0.954	0.947	0.964	0.956	35	23	58
8	Negative impacts of climate change on health	`health_neg_impact`	0.386	0.340	0.364	0.629	0.514	0.571	0.485	0.349	0.416	22	18	40
9	Health co-benefits of climate action	`health_cobenefit`	0.252	0.103	0.186	0.257	0.114	0.186	0.301	0.023	0.177	4	4	8
Security Frame														
10	**Security Frame (Primary Category)**	`security_frame`	0.825	0.730	0.782	0.951	0.939	0.945	0.955	0.950	0.952	30	19	49

Continued on next page

Table B5 – *Continued from previous page*

#	Category	Code	F1 Macro			F1 Micro			F1 Weighted			Support		
			EN	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**
11	Presence of climate refugees	`security_refugees`	0.476	0.417	0.481	0.545	0.429	0.488	0.468	0.476	0.459	21	6	27
12	Conflict	`security_conflict`	0.358	0.286	0.324	0.364	0.286	0.326	0.398	0.286	0.346	7	6	13
13	Post-disaster military assistance	`security_military`	0.000	0.045	0.045	0.000	0.048	0.048	0.000	0.004	0.004	0	2	2
14	Disruption of military operations	`security_disruption`	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0	2	2
Justice Frame														
15	**Justice Frame (Primary Category)**	`justice_frame`	0.832	0.880	0.857	0.917	0.937	0.927	0.921	0.939	0.930	62	74	136
16	Winners and losers of climate action	`justice_winners`	0.654	0.737	0.708	0.765	0.762	0.764	0.802	0.775	0.786	10	22	32
17	Differentiated responsibility	`justice_responsibility`	0.900	0.823	0.860	0.926	0.833	0.879	0.927	0.838	0.883	19	27	46
18	Unequal vulnerability to climate change	`justice_vulnerability`	0.786	0.894	0.845	0.852	0.917	0.885	0.858	0.918	0.888	16	21	37
19	Unequal access to climate action	`justice_access`	0.475	0.796	0.651	0.481	0.833	0.661	0.456	0.835	0.674	27	23	50
20	Intergenerational justice	`justice_intergen`	0.924	0.952	0.942	0.963	0.964	0.964	0.965	0.965	0.965	10	19	29
Political Frame														
21	**Political Frame (Primary Category)**	`political_frame`	0.812	0.842	0.829	0.827	0.844	0.836	0.829	0.845	0.837	167	217	384
22	Policy action	`pol_action`	0.802	0.722	0.756	0.914	0.864	0.886	0.909	0.854	0.879	26	38	64
23	Political debate	`pol_debate`	0.469	0.650	0.564	0.694	0.820	0.763	0.595	0.781	0.699	127	175	302
24	Political positioning	`pol_position`	0.752	0.808	0.782	0.801	0.912	0.862	0.821	0.920	0.877	35	24	59
25	Public opinion data	`pol_opinion`	0.877	0.910	0.894	0.962	0.974	0.969	0.966	0.976	0.971	12	15	27
Scientific Frame														
26	**Scientific Frame (Primary Category)**	`scientific_frame`	0.891	0.890	0.890	0.951	0.949	0.950	0.948	0.947	0.947	74	75	149

Continued on next page

Table B5 – *Continued from previous page*

#	Category	Code	F1 Macro			F1 Micro			F1 Weighted			Support		
			EN	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**
27	Scientific debate	`sci_debate`	0.962	0.867	0.911	0.962	0.869	0.912	0.962	0.869	0.912	26	26	52
28	Popularisation or scientific discovery	`sci_discovery`	0.884	0.797	0.838	0.885	0.803	0.841	0.886	0.803	0.842	31	36	67
29	Questioning of climate science	`sci_skepticism`	0.783	0.768	0.775	0.846	0.836	0.841	0.860	0.857	0.858	9	9	18
30	Defense of climate science	`sci_defense`	0.769	0.460	0.682	0.846	0.852	0.850	0.862	0.785	0.843	8	9	17
Environmental Frame														
31	**Environmental Frame (Primary Category)**	`environmental_frame`	0.866	0.875	0.871	0.967	0.967	0.967	0.969	0.968	0.969	28	32	60
32	Loss of natural environments	`env_habitat`	0.804	0.573	0.721	0.806	0.707	0.753	0.802	0.637	0.730	16	26	42
33	Loss of fauna and flora	`env_species`	0.714	0.692	0.703	0.722	0.707	0.714	0.717	0.694	0.705	19	21	40
Cultural Frame														
34	**Cultural Frame (Primary Category)**	`cultural_frame`	0.802	0.581	0.708	0.923	0.880	0.901	0.931	0.920	0.921	41	11	52
35	Artistic representation	`cult_art`	0.834	0.469	0.659	0.851	0.559	0.704	0.858	0.649	0.739	18	6	24
36	Difficulty to host cultural or sports events	`cult_event_impact`	0.631	0.354	0.484	0.731	0.456	0.593	0.777	0.595	0.686	8	2	10
37	Loss of indigenous practices	`cult_indigenous`	0.824	0.415	0.595	0.910	0.618	0.763	0.922	0.749	0.825	7	1	8
38	Carbon footprint of the cultural and sports sectors	`cult_footprint`	0.702	0.014	0.351	0.925	0.015	0.467	0.945	0.000	0.613	2	1	3
PRIMARY CATEGORIES														
Actors/Messengers														
39	**Presence of Messengers (Primary Category)**	`messenger`	0.955	0.967	0.961	0.957	0.969	0.963	0.957	0.969	0.963	301	312	613
40	Health expert	`msg_health`	0.862	0.951	0.907	0.964	0.987	0.975	0.966	0.987	0.976	19	21	40

Continued on next page

Table B5 – *Continued from previous page*

#	Category	Code	F1 Macro			F1 Micro			F1 Weighted			Support		
			EN	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**
41	Economic expert	`msg_economic`	0.847	0.895	0.874	0.954	0.961	0.957	0.949	0.963	0.957	30	28	58
42	Security expert	`msg_security`	0.728	0.813	0.787	0.970	0.958	0.964	0.973	0.962	0.968	7	14	21
43	Legal expert	`msg_legal`	0.861	0.927	0.914	0.990	0.980	0.985	0.989	0.982	0.986	7	19	26
44	Cultural or Sport expert	`msg_cultural`	0.853	0.825	0.843	0.954	0.967	0.961	0.954	0.969	0.961	26	13	39
45	Natural scientist	`msg_scientist`	0.920	0.896	0.907	0.960	0.941	0.951	0.960	0.944	0.952	44	46	90
46	Social scientist	`msg_social`	0.938	0.948	0.944	0.987	0.984	0.985	0.987	0.984	0.986	17	24	41
47	Activist	`msg_activist`	0.934	0.903	0.918	0.977	0.964	0.970	0.976	0.963	0.970	31	33	64
48	Public official	`msg_official`	0.927	0.930	0.929	0.934	0.935	0.934	0.934	0.935	0.934	104	113	217
Events														
49	**Presence of Events (Primary Category)**	`event`	0.881	0.912	0.897	0.886	0.913	0.900	0.885	0.913	0.900	201	222	423
50	Extreme meteorological event	`evt_weather`	0.887	0.891	0.889	0.935	0.955	0.946	0.938	0.958	0.949	29	21	50
51	Meeting	`evt_meeting`	0.855	0.891	0.877	0.908	0.914	0.911	0.914	0.918	0.916	30	50	80
52	Publication	`evt_publication`	0.950	0.962	0.958	0.973	0.968	0.970	0.973	0.969	0.971	30	62	92
53	Election	`evt_election`	0.911	0.895	0.903	0.962	0.964	0.963	0.960	0.963	0.962	25	22	47
54	New policy	`evt_policy`	0.821	0.837	0.829	0.886	0.914	0.901	0.895	0.915	0.906	29	33	62
55	Judiciary decision	`evt_judiciary`	0.628	0.939	0.808	0.903	0.982	0.946	0.927	0.981	0.951	6	19	25
56	Cultural or Sports event	`evt_cultural`	0.906	0.857	0.881	0.973	0.964	0.968	0.971	0.967	0.968	17	12	29
57	Protest	`evt_protest`	0.982	0.909	0.943	0.995	0.973	0.983	0.995	0.973	0.983	16	18	34
Solutions														
58	**Presence of Solutions (Primary Category)**	`solution`	0.891	0.909	0.900	0.921	0.931	0.926	0.922	0.930	0.926	114	133	247
59	Mitigation strategy	`sol_mitigation`	0.872	0.801	0.839	0.884	0.839	0.861	0.882	0.850	0.864	76	97	173
60	Adaptation strategy	`sol_adaptation`	0.889	0.903	0.895	0.934	0.952	0.943	0.934	0.954	0.944	22	16	38
EMOTIONAL TONE														
61	Emotion: positive	`tone_positive`	0.712	0.692	0.702	0.902	0.892	0.897	0.910	0.904	0.907	38	37	75
62	Emotion: negative	`tone_negative`	0.793	0.804	0.799	0.807	0.813	0.810	0.814	0.819	0.816	151	165	316
63	Emotion: neutral	`tone_neutral`	0.756	0.798	0.777	0.764	0.809	0.787	0.765	0.807	0.787	295	306	601

Continued on next page

Table B5 – *Continued from previous page*

#	Category	Code	F1 Macro			F1 Micro			F1 Weighted			Support		
			EN	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**	**EN**	**FR**	**ALL**
GEOGRAPHIC FOCUS														
64	Mention of Canada	`canada`	0.986	0.978	0.982	0.986	0.978	0.982	0.986	0.978	0.982	235	255	490
URGENCY TO ACT														
65	Urgency to act	`urgency`	0.874	0.835	0.853	0.976	0.965	0.970	0.977	0.967	0.972	23	25	48
OVERALL PERFORMANCE														
		All Categories	**0.869**	**0.864**	**0.866**	**0.909**	**0.902**	**0.905**	**0.911**	**0.905**	**0.908**	**3069**	**3332**	**6401**

Note: Metrics represent macro-averaged scores across positive and negative classes for each category. Support indicates the number of positive examples in the validation set (500 sentences per language, 1,000 total). EN = English, FR = French, ALL = Combined. Categories are organized by their hierarchical grouping as implemented in the annotation pipeline.

Table B6: Database-wide distribution of annotation categories across 9.2 million sentences

#	Category	Code	Count			Proportion (%)		
			EN	FR	ALL	EN	FR	ALL
THEMATIC FRAMES								
Economic Frame								
1	**Economic Frame (Primary Category)**	`economic_frame`	1,232,910	185,032	1,417,942	15.94	12.64	15.41
2	Negative impacts of climate change on the economy	`eco_neg_impact`	193,068	30,670	223,738	2.50	2.10	2.43
3	Positive impacts of climate change on the economy	`eco_pos_impact`	1,153	1,949	3,102	0.01	0.13	0.03
4	Economic disadvantages of climate action	`eco_cost`	516,374	40,259	556,633	6.68	2.75	6.05
5	Economic benefits of climate action	`eco_benefit`	75,145	30,266	105,411	0.97	2.07	1.15
6	Carbon footprint of the economic sector	`eco_footprint`	292,909	61,974	354,883	3.79	4.23	3.86
Health Frame								
7	**Health Frame (Primary Category)**	`health_frame`	106,679	21,164	127,843	1.38	1.45	1.39
8	Negative impacts of climate change on health	`health_neg_impact`	106,679	21,164	127,843	1.38	1.45	1.39
9	Health co-benefits of climate action	`health_cobenefit`	63,774	19,595	83,369	0.82	1.34	0.91
Security Frame								
10	**Security Frame (Primary Category)**	`security_frame`	32,925	10,255	43,180	0.43	0.70	0.47

Continued on next page

Table B6 – *Continued from previous page*

#	Category	Code	Count EN	Count FR	Count ALL	Proportion (%) EN	Proportion (%) FR	Proportion (%) ALL
11	Presence of climate refugees	`security_refugees`	21,384	4,248	25,632	0.28	0.29	0.28
12	Conflict	`security_conflict`	20,405	9,015	29,420	0.26	0.62	0.32
13	Post-disaster military assistance	`security_military`	0	10,253	10,253	0.00	0.70	0.11
14	Disruption of military operations	`security_disruption`	0	4	4	0.00	0.00	0.00
	Justice Frame							
15	**Justice Frame (Primary Category)**	`justice_frame`	276,401	20,938	297,339	3.57	1.43	3.23
16	Winners and losers of climate action	`justice_winners`	45,550	8,905	54,455	0.59	0.61	0.59
17	Differentiated responsibility	`justice_responsibility`	59,556	9,084	68,640	0.77	0.62	0.75
18	Unequal vulnerability to climate change	`justice_vulnerability`	13,194	2,188	15,382	0.17	0.15	0.17
19	Unequal access to climate action	`justice_access`	215,212	4,791	220,003	2.78	0.33	2.39
20	Intergenerational justice	`justice_intergen`	26,797	1,160	27,957	0.35	0.08	0.30
	Political Frame							
21	**Political Frame (Primary Category)**	`political_frame`	2,408,280	409,831	2,818,111	31.13	28.00	30.64
22	Policy action	`pol_action`	204,684	18,833	223,517	2.65	1.29	2.43
23	Political debate	`pol_debate`	2,357,345	401,652	2,758,997	30.47	27.44	29.99
24	Political positioning	`pol_position`	860,176	49,748	909,924	11.12	3.40	9.89
25	Public opinion data	`pol_opinion`	46,825	8,186	55,011	0.61	0.56	0.60
	Scientific Frame							

Continued on next page

Table B6 – *Continued from previous page*

#	Category	Code	Count			Proportion (%)		
			EN	FR	ALL	EN	FR	ALL
26	**Scientific Frame (Primary Category)**	`scientific_frame`	516,670	75,144	591,814	6.68	5.13	6.43
27	Scientific debate	`sci_debate`	212,607	14,929	227,536	2.75	1.02	2.47
28	Popularisation or scientific discovery	`sci_discovery`	344,877	62,495	407,372	4.46	4.27	4.43
29	Questioning of climate science	`sci_skepticism`	64,736	2,065	66,801	0.84	0.14	0.73
30	Defense of climate science	`sci_defense`	48,511	0	48,511	0.63	0.00	0.53
Environmental Frame								
31	**Environmental Frame (Primary Category)**	`environmental_frame`	206,531	74,554	281,085	2.67	5.09	3.06
32	Loss of natural environments	`env_habitat`	107,509	62,541	170,050	1.39	4.27	1.85
33	Loss of fauna and flora	`env_species`	118,580	49,797	168,377	1.53	3.40	1.83
Cultural Frame								
34	**Cultural Frame (Primary Category)**	`cultural_frame`	290,437	82,465	372,902	3.75	5.63	4.05
35	Artistic representation	`cult_art`	202,421	44,568	246,989	2.62	3.05	2.68
36	Difficulty to host cultural or sports events	`cult_event_impact`	22,993	34,855	57,848	0.30	2.38	0.63
37	Loss of indigenous practices	`cult_indigenous`	7,200	18,778	25,978	0.09	1.28	0.28
38	Carbon footprint of the cultural and sports sectors	`cult_footprint`	13	82,382	82,395	0.00	5.63	0.90
PRIMARY CATEGORIES								
Actors/Messengers								
39	**Presence of Messengers (Primary Category)**	`messenger`	3,843,269	644,020	4,487,289	49.68	44.00	48.78

Continued on next page

Table B6 – *Continued from previous page*

#	Category	Code	Count			Proportion (%)		
			EN	FR	ALL	EN	FR	ALL
40	Health expert	`msg_health`	47,074	7,428	54,502	0.61	0.51	0.59
41	Economic expert	`msg_economic`	215,924	73,343	289,267	2.79	5.01	3.14
42	Security expert	`msg_security`	19	8,916	8,935	0.00	0.61	0.10
43	Legal expert	`msg_legal`	7	5,025	5,032	0.00	0.34	0.05
44	Cultural or Sport expert	`msg_cultural`	123,890	18,037	141,927	1.60	1.23	1.54
45	Natural scientist	`msg_scientist`	279,621	70,656	350,277	3.61	4.83	3.81
46	Social scientist	`msg_social`	32,227	19,609	51,836	0.42	1.34	0.56
47	Activist	`msg_activist`	96,990	25,189	122,179	1.25	1.72	1.33
48	Public official	`msg_official`	1,285,694	223,649	1,509,343	16.62	15.28	16.41
Events								
49	**Presence of Events (Primary Category)**	`event`	1,419,146	277,560	1,696,706	18.35	18.96	18.44
50	Extreme meteorological event	`evt_weather`	199,964	38,746	238,710	2.59	2.65	2.59
51	Meeting	`evt_meeting`	396,195	94,979	491,174	5.12	6.49	5.34
52	Publication	`evt_publication`	234,021	72,729	306,750	3.03	4.97	3.33
53	Election	`evt_election`	64,641	28,706	93,347	0.84	1.96	1.01
54	New policy	`evt_policy`	379,574	42,444	422,018	4.91	2.90	4.59
55	Judiciary decision	`evt_judiciary`	37,007	6,285	43,292	0.48	0.43	0.47
56	Cultural or Sports event	`evt_cultural`	964	22,998	23,962	0.01	1.57	0.26
57	Protest	`evt_protest`	7,728	5,293	13,021	0.10	0.36	0.14
Solutions								
58	**Presence of Solutions (Primary Category)**	`solution`	1,818,537	307,710	2,126,247	23.51	21.02	23.11
59	Mitigation strategy	`sol_mitigation`	1,258,602	201,352	1,459,954	16.27	13.76	15.87
60	Adaptation strategy	`sol_adaptation`	48,944	18,119	67,063	0.63	1.24	0.73

Continued on next page

Table B6 – *Continued from previous page*

#	Category	Code	Count			Proportion (%)		
			EN	**FR**	**ALL**	**EN**	**FR**	**ALL**
EMOTIONAL TONE								
61	Emotion: positive	`tone_positive`	633,227	184,697	817,924	8.19	12.62	8.89
62	Emotion: negative	`tone_negative`	2,971,135	513,870	3,485,005	38.41	35.11	37.88
63	Emotion: neutral	`tone_neutral`	4,872,522	1,057,425	5,929,947	62.99	72.25	64.46
GEOGRAPHIC FOCUS								
64	Mention of Canada	`canada`	3,383,427	573,263	3,956,690	43.74	39.17	43.01
URGENCY TO ACT								
65	Urgency to act	`urgency`	160,053	35,784	195,837	2.07	2.44	2.13
Total Sentences			**7,735,377**	**1,463,581**	**9,198,958**	**100.00**	**100.00**	**100.00**

Note: Counts represent the number of sentences in the CCF database annotated with each category. Proportions show the percentage of sentences containing each annotation. Categories are organized by their hierarchical grouping as implemented in the annotation pipeline. EN = English corpus, FR = French corpus, ALL = Combined corpus.

References

[1] Dietram A. Scheufele and David Tewksbury. "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models". en. In: *Journal of Communication* 57.1 (Mar. 2007). Publisher: John Wiley & Sons, Ltd, pp. 9–20. ISSN: 1460-2466. DOI: 10.1111/j.0021-9916.2007.00326.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.0021-9916.2007.00326.x (visited on 12/05/2025).

[2] Stefaan Walgrave and Peter Van Aelst. "The Contingency of the Mass Media's Political Agenda Setting Power: Toward a Preliminary Theory". en. In: *Journal of Communication* 56.1 (Mar. 2006), pp. 88–109. ISSN: 0021-9916, 1460-2466. DOI: 10.1111/j.1460-2466.2006.00005.x. URL: https://academic.oup.com/joc/article/56/1/88-109/4102573 (visited on 12/05/2025).

[3] Stanley Cohen. *Folk Devils and Moral Panics - The Creation of the Mods and Rockers*. London: MacGibbon and Kee Ltd, 1972. ISBN: 978-0-203-82825-0.

[4] William Hoynes. "Media and the Construction of Social Problems". In: *The Cambridge Handbook of Social Problems*. Ed. by A. Javier Treviño. Vol. 1. Cambridge: Cambridge University Press, 2018, pp. 477–496. ISBN: 978-1-108-42616-9. DOI: 10.1017/9781108656184.027. URL: https://www.cambridge.org/core/books/cambridge-handbook-of-social-problems/media-and-the-construction-of-social-problems/70A9AFB47F08651CAB1B944034BB18FA (visited on 12/05/2025).

[5] Yiwei Luo, Dallas Card, and Dan Jurafsky. *Detecting Stance in Media on Global Warming*. arXiv:2010.15149 [cs]. Jan. 2021. DOI: 10.48550/arXiv.2010.15149. URL: http://arxiv.org/abs/2010.15149 (visited on 12/05/2025).

[6] Paul Meddeb et al. "Counteracting French Fake News on Climate Change Using Language Models". en. In: *Sustainability* 14.18 (Sept. 2022). Company: Multidisciplinary Digital Publishing Institute Distributor: Multidisciplinary Digital Publishing Institute Institution: Multidisciplinary Digital Publishing Institute Label: Multidisciplinary Digital Publishing Institute Publisher: publisher. ISSN: 2071-1050. DOI: 10.3390/su141811724. URL: https://www.mdpi.com/2071-1050/14/18/11724 (visited on 12/05/2025).

[7] Newspapers Canada. *Circulation Report: Daily Newspapers*. Tech. rep. Newspapers Canada, 2015. URL: https://nmc-mic.ca/wp-content/uploads/2016/06/2015-Daily-Newspaper-Circulation-Report-REPORT_FINAL.pdf (visited on 12/05/2025).

[8] Robert M. Entman. "Framing: Toward Clarification of a Fractured Paradigm". en. In: *Journal of Communication* 43.4 (Dec. 1993), pp. 51–58. ISSN: 0021-9916, 1460-2466. DOI: 10.1111/j.1460-2466.1993.tb01304.x. URL: https://academic.oup.com/joc/article/43/4/51-58/4160153 (visited on 12/05/2025).

[9] Reiner Grundmann. "Using large text news archives for the analysis of climate change discourse: some methodological observations". In: *Journal of Risk Research* 25.3 (Mar. 2022). Publisher: Routledge _eprint: https://doi.org/10.1080/13669877.2021.1894471, pp. 395–406. ISSN: 1366-9877. DOI: 10.1080/13669877.2021.1894471. URL: https://doi.org/10.1080/13669877.2021.1894471 (visited on 12/05/2025).

[10] Reiner Grundmann. “Climate change as a wicked social problem”. en. In: *Nature Geoscience* 9.8 (Aug. 2016). Publisher: Nature Publishing Group, pp. 562–563. ISSN: 1752-0908. DOI: 10.1038/ngeo2780. URL: https://www.nature.com/articles/ngeo2780 (visited on 12/05/2025).

[11] Manfred Stede and Ronny Patz. “The Climate Change Debate and Natural Language Processing”. In: *Proceedings of the 1st Workshop on NLP for Positive Impact*. Ed. by Anjalie Field et al. Online: Association for Computational Linguistics, Aug. 2021, pp. 8–18. DOI: 10.18653/v1/2021.nlp4posimpact-1.2. URL: https://aclanthology.org/2021.nlp4posimpact-1.2/ (visited on 12/05/2025).

[12] James Ford and Diana King. “Coverage and framing of climate change adaptation in the media: A review of influential North American newspapers during 1993–2013”. In: *Environmental Science & Policy* 48 (Apr. 2015), pp. 137–146. ISSN: 1462-9011. DOI: 10.1016/j.envsci.2014.12.003. URL: https://www.sciencedirect.com/science/article/pii/S1462901114002317 (visited on 12/05/2025).

[13] Craig Trumbo. “Constructing climate change: claims and frames in US news coverage of an environmental issue”. en. In: *Public Understanding of Science* 5.3 (July 1996), p. 269. ISSN: 0963-6625. DOI: 10.1088/0963-6625/5/3/006. URL: https://doi.org/10.1088/0963-6625/5/3/006 (visited on 12/05/2025).

[14] Maxwell T. Boykoff. *Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change*. en. Google-Books-ID: zsdRZR88tdkC. Cambridge University Press, Sept. 2011. ISBN: 978-1-139-50179-8.

[15] Anthony Downs. “Up and down with ecology—the issue attention cycle”. In: *Public Interest* 28 (1972), pp. 38–50. URL: https://fbaum.unc.edu/teaching/articles/Downs_Public_Interest_1972.pdf (visited on 12/05/2025).

[16] Xinsheng Liu, Eric Lindquist, and Arnold Vedlitz. “Explaining Media and Congressional Attention to Global Climate Change, 1969-2005: An Empirical Test of Agenda-Setting Theory”. en. In: *Political Research Quarterly* 64.2 (2011), pp. 405–419. DOI: 10.1177/1065912909346744. URL: https://journals.sagepub.com/doi/epdf/10.1177/1065912909346744 (visited on 12/05/2025).

[17] Susanne C. Moser. “Reflections on climate change communication research and practice in the second decade of the 21st century: what more is there to say?” en. In: *Wiley Interdisciplinary Reviews: Climate Change* 7.3 (May 2016). Publisher: John Wiley & Sons, Ltd, pp. 345–369. ISSN: 1757-7799. DOI: 10.1002/wcc.403. URL: https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.403 (visited on 12/06/2025).

[18] Sefat Salama and Khalil Aboukoura. “Role of Emotions in Climate Change Communication”. en. In: *Handbook of Climate Change Communication: Vol. 1: Theory of Climate Change Communication*. Ed. by Walter Leal Filho et al. Cham: Springer International Publishing, 2018, pp. 137–150. ISBN: 978-3-319-69838-0. DOI: 10.1007/978-3-319-69838-0_9. URL: https://doi.org/10.1007/978-3-319-69838-0_9 (visited on 12/05/2025).

[19] Robin L. Nabi, Abel Gustafson, and Risa Jensen. "Framing Climate Change: Exploring the Role of Emotion in Generating Advocacy Behavior". EN. In: *Science Communication* 40.4 (Aug. 2018). Publisher: SAGE Publications Inc, pp. 442–468. ISSN: 1075-5470. DOI: 10.1177/1075547018776019. URL: https://doi.org/10.1177/1075547018776019 (visited on 12/05/2025).

[20] Salomé Do, Étienne Ollion, and Rubing Shen. "The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy". en. In: *Sociological Methods & Research* (Dec. 2022). Publisher: SAGE Publications Inc, p. 00491241221134526. ISSN: 0049-1241. DOI: 10.1177/00491241221134526. URL: https://doi.org/10.1177/00491241221134526 (visited on 04/15/2024).

[21] Antoine Lemor. *antoinelemor/AugmentedSocialScientistFork*. original-date: 2025-04-26T21:39:03Z. June 2025. URL: https://github.com/antoinelemor/AugmentedSocialScientistFo (visited on 06/19/2025).

[22] Marco Toledo Bastos. "Shares, Pins, and Tweets: News readership from daily papers to social media". In: *Journalism Studies* 16.3 (May 2015). Publisher: Routledge _eprint: https://doi.org/10.1080/1461670X.2014.891857, pp. 305–325. ISSN: 1461-670X. DOI: 10.1080/1461670X.2014.891857. URL: https://doi.org/10.1080/1461670X.2014.891857 (visited on 12/05/2025).

[23] Andreu Casero-Ripolles. *Impact of COVID-19 on the Media System. Communicative and Democratic Consequences of News Consumption during the Outbreak*. en. SSRN Scholarly Paper. Rochester, NY, Apr. 2020. URL: https://papers.ssrn.com/abstract=3594133 (visited on 12/05/2025).

[24] N. Lulham et al. *Canada in a changing climate: synthesis report*. English. Tech. rep. Natural Resources Canada, Dec. 2023. URL: https://ostrnrcan-dostrncan.canada.ca/handle/1845/271933 (visited on 12/05/2025).

[25] Natural Resources Canada. *Energy Fact Book, 2025-2026*. Tech. rep. Natural Resources Canada, 2025. URL: https://energy-information.canada.ca/sites/default/files/2025-11/energy-fact-book-2025-2026.pdf (visited on 12/05/2025).

[26] Kathryn Harrison. "The Struggle of Ideas and Self-Interest in Canadian Climate Policy". en. In: *Global Commons, Domestic Decisions: The Comparative Politics of Climate Change*. Ed. by Kathryn Harrison and Lisa McIntosh Sundstrom. Google-Books-ID: VsTxCwAAQBAJ. MIT Press, July 2010, p. 336. ISBN: 978-0-262-28887-3.

[27] C. Hatch et al. *What Do Canadians Really Think About Climate Change? A Summary of Public Opinion Research for Communicators*. Tech. rep. Re.Climate, 2025. URL: https://reclimate.ca/wp-content/uploads/2025/06/Re.Climate-Public-Opinion-Summary-2025-Report.pdf (visited on 12/05/2025).

[28] Nathan Young and Eric Dugas. "Comparing Climate Change Coverage in Canadian English- and French-Language Print Media: Environmental Values, Media Cultures, and the Narration of Global Warming". In: *The Canadian Journal of Sociology / Cahiers canadiens de sociologie* 37.1 (2012). Publisher: Canadian Journal of Sociology, pp. 25–54. ISSN: 0318-6431. URL: https://www.jstor.org/stable/canajsocicahican.37.1.25 (visited on 12/05/2025).

[29] Nia King et al. "How do Canadian media report climate change impacts on health? A newspaper review". en. In: *Climatic Change* 152.3 (Mar. 2019), pp. 581–596. ISSN: 1573-1480. DOI: 10.1007/s10584-018-2311-2. URL: https://doi.org/10.1007/s10584-018-2311-2 (visited on 12/05/2025).

[30] Pradip Swarnakar and Ashutosh Modi. *NLP for Climate Policy: Creating a Knowledge Platform for Holistic and Effective Climate Action*. arXiv:2105.05621 [cs]. May 2021. DOI: 10.48550/arXiv.2105.05621. URL: http://arxiv.org/abs/2105.05621 (visited on 12/05/2025).

[31] Mike S. Schäfer and Valerie Hase. "Computational methods for the analysis of climate change communication: Towards an integrative and reflexive approach". en. In: *Wiley Interdisciplinary Reviews: Climate Change* 14.2 (Mar. 2023). Publisher: John Wiley & Sons, Ltd, e806. ISSN: 1757-7799. DOI: 10.1002/wcc.806. URL: https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.806 (visited on 12/05/2025).

[32] Simon David Hirsbrunner. "Computational methods for climate change frame analysis: Techniques, critiques, and cautious ways forward". en. In: *Wiley Interdisciplinary Reviews: Climate Change* 15.5 (Sept. 2024). Publisher: John Wiley & Sons, Ltd, e902. ISSN: 1757-7799. DOI: 10.1002/wcc.902. URL: https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.902 (visited on 12/05/2025).

[33] Tom Nicholls and Pepper D. Culpepper. "Computational Identification of Media Frames: Strengths, Weaknesses, and Opportunities". In: *Computational Political Communication*. Ed. by Yannis Theocharis and Andreas Jungherr. Num Pages: 23. Routledge, 2024.

[34] Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. "Media Framing: A typology and Survey of Computational Approaches Across Disciplines". In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Lun-Wei Ku, Andre Martins, and Vivek Srikumar. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 15407–15428. DOI: 10.18653/v1/2024.acl-long.822. URL: https://aclanthology.org/2024.acl-long.822/ (visited on 12/05/2025).

[35] Regina G. Lawrence. "Framing Obesity: The Evolution of News Discourse on a Public Health Issue". EN. In: *Harvard International Journal of Press/Politics* 9.3 (July 2004). Publisher: SAGE Publications, pp. 56–75. ISSN: 1081-180X. DOI: 10.1177/1081180X04266581. URL: https://doi.org/10.1177/1081180X04266581 (visited on 12/05/2025).

[36] Jennifer Good. "The Framing of Climate Change in Canadian, American, and International Newspapers: A Media Propaganda Model Analysis". In: *Canadian Journal of Communication* 33.2 (June 2008). Publisher: University of Toronto Press, pp. 233–256. ISSN: 0705-3657. DOI: 10.22230/cjc.2008v33n2a2017. URL: https://cjc.utppublishing.com/doi/full/10.22230/cjc.2008v33n2a2017 (visited on 12/05/2025).

[37] Dan Rowe. “Comparing Newspaper Coverage of Climate Change During Election Campaigns in the United States, Canada and Australia”. Anglais. ISBN: 9781267179050. Ph.D. 2011. URL: https://www.proquest.com/docview/922669062/abstract/2CFD056A38334D6EPQ/1 (visited on 12/05/2025).

[38] Nathan Young and Eric Dugas. “Representations of Climate Change in Canadian National Print Media: The Banalization of Global Warming”. en. In: *Canadian Review of Sociology/Revue canadienne de sociologie* 48.1 (Feb. 2011). Publisher: John Wiley & Sons, Ltd, pp. 1–22. ISSN: 1755-618X. DOI: 10.1111/j.1755-618X.2011.01247.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1755-618X.2011.01247.x (visited on 12/05/2025).

[39] Darryn Anne DiFrancesco and Nathan Young. “Seeing climate change: the visual construction of global warming in Canadian national print media”. EN. In: *cultural geographies* 18.4 (Oct. 2011). Publisher: SAGE Publications Ltd, pp. 517–536. ISSN: 1474-4740. DOI: 10.1177/1474474010382072. URL: https://doi.org/10.1177/1474474010382072 (visited on 12/05/2025).

[40] Katrina Ahchong and Rachel Dodds. “Anthropogenic climate change coverage in two Canadian newspapers, the *Toronto Star* and the *Globe and Mail*, from 1988 to 2007”. In: *Environmental Science & Policy* 15.1 (Jan. 2012), pp. 48–59. ISSN: 1462-9011. DOI: 10.1016/j.envsci.2011.09.008. URL: https://www.sciencedirect.com/science/article/pii/S146290111100150X (visited on 12/05/2025).

[41] Mark Stoddart and David Tindall. “Canadian news media and the cultural dynamics of multilevel climate governance”. In: *Environmental Politics* 24.3 (May 2015). Publisher: Routledge _eprint: https://doi.org/10.1080/09644016.2015.1008249, pp. 401–422. ISSN: 0964-4016. DOI: 10.1080/09644016.2015.1008249. URL: https://doi.org/10.1080/09644016.2015.1008249 (visited on 12/05/2025).

[42] Mark Stoddart and Jillian Smith. “The Endangered Arctic, the Arctic as Resource Frontier: Canadian News Media Narratives of Climate Change and the North”. en. In: *Canadian Review of Sociology/Revue canadienne de sociologie* 53.3 (Aug. 2016). Publisher: John Wiley & Sons, Ltd, pp. 316–336. ISSN: 1755-618X. DOI: 10.1111/cars.12111. URL: https://onlinelibrary.wiley.com/doi/10.1111/cars.12111 (visited on 12/05/2025).

[43] Mark Stoddart et al. “Media Access and Political Efficacy in the Eco-politics of Climate Change: Canadian National News and Mediated Policy Networks”. In: *Environmental Communication* 11.3 (May 2017). Publisher: Routledge _eprint: https://doi.org/10.1080/17524032.201 pp. 386–400. ISSN: 1752-4032. DOI: 10.1080/17524032.2016.1275731. URL: https://doi.org/10.1080/17524032.2016.1275731 (visited on 12/05/2025).

[44] Ralf Barkemeyer et al. “Media coverage of climate change: An international comparison”. EN. In: *Environment and Planning C: Politics and Space* 35.6 (Sept. 2017). Publisher: SAGE Publications Ltd STM, pp. 1029–1054. ISSN: 2399-6544. DOI: 10.1177/0263774X16680818. URL: https://doi.org/10.1177/0263774X16680818 (visited on 12/05/2025).

[45] Alizee Pillod. “Reframing climate change as a public health issue : a Canadian case study, 2008-2020”. fr. PhD thesis. July 2021. URL: http://hdl.handle.net/1866/26126 (visited on 12/05/2025).

[46] Oliver Brandt. “News Media Framing of Climate Change in Canada: A Provincial Analysis of 2015–2018”. Anglais. ISBN: 9798496553797. MA thesis. 2021. URL: https://www.proquest.com/docview/2607337799/abstract/16DEF7A383FD4B8DPQ/1 (visited on 12/05/2025).

[47] Valerie Hase et al. “Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006–2018)”. In: *Global Environmental Change* 70 (Sept. 2021), p. 102353. ISSN: 0959-3780. DOI: 10.1016/j.gloenvcha.2021.102353. URL: https://www.sciencedirect.com/science/article/pii/S0959378021001321 (visited on 12/05/2025).

[48] Sonya Sachdeva and Sarah McCaffrey. “Themes and patterns in print media coverage of wildfires in the USA, Canada and Australia: 1986–2016”. In: *International Journal of Wildland Fire* 31.12 (Oct. 2022), pp. 1089–1102. ISSN: 1049-8001. DOI: 10.1071/WF22174. URL: https://doi.org/10.1071/WF22174 (visited on 12/05/2025).

[49] Mark Stoddart and Yixi Yang. “What are the roles of regional and local climate governance discourse and actors? Mediated climate change policy networks in Atlantic Canada”. en. In: *Review of Policy Research* 40.6 (Nov. 2023). Publisher: John Wiley & Sons, Ltd, pp. 1144–1168. ISSN: 1541-1338. DOI: 10.1111/ropr.12510. URL: https://onlinelibrary.wiley.com/doi/10.1111/ropr.12510 (visited on 12/05/2025).

[50] Kaiping Chen et al. “How Climate Movement Actors and News Media Frame Climate Change and Strike: Evidence from Analyzing Twitter and News Media Discourse from 2018 to 2021”. EN. In: *The International Journal of Press/Politics* 28.2 (Apr. 2023). Publisher: SAGE Publications Inc, pp. 384–413. ISSN: 1940-1612. DOI: 10.1177/19401612221106405. URL: https://doi.org/10.1177/19401612221106405 (visited on 12/05/2025).