

EUROPEAN POLITICAL SCIENCE SOCIETY ANNUAL CONFERENCE, BELFAST, JUNE 18-20, 2026

CCF Database

A machine-learning-preprocessed media corpus of 266,000 climate articles

65 sentence-level annotation categories · 20 newspapers · 1978-2024 · EN + FR

Antoine Lemor¹, Alizée Pillod², Matthew Taylor²

¹Université de Sherbrooke, ²Université de Montréal

CAPP · CIRST · RFICS



data.ccf-project.ca

EPSS2026

Outline

1

Why CCF

from a single article to a 35-year transformation

2

Method

corpus · annotation pipeline · 65 categories

3

Applications

frame evolution · regional patterns · media
cascades

PART I

Why CCF

The puzzle of climate discourse, and what we are missing

Two summers, two ways of telling the same crisis

Toronto Star

SUMMER 1988

Summer of '88 could be hint of the future. Scientists fear warming trend will

Globe and Mail

SUMMER 2023

*Poillievre seizes the moment on Atlantic Canada carbon-tax fears. (...) The
Conservative leader said...*

Same type of events, but different dominant actors and frames.

RESEARCH QUESTION

If two summer reports can read so differently, what would it take to measure climate discourse, actors and frames across every outlet, every region, every year?

Climate framing matters, but our measurement tools have not kept up

How issues are framed in the news shapes **what publics believe, what coalitions form, and what policies become thinkable.**

ENTMAN 1993 · BOYKOFF & BOYKOFF 2007 · NISBET 2009

Canadian climate discourse is fragmented, bilingual, regional. Existing datasets cover **short windows, single outlets, English only.**

YOUNG & DUGAS 2011 · STODDART ET AL. 2017 · BOURSIER & LEMOR 2024

Climate communication research has documented framing dynamics through « *hand coding* » of small samples, or « *topic modelling* » of large but unsupervised corpora.

STECULA & MERKLEY 2019 · BOHR 2020 · TVINNEREIM & FLØTTUM 2015

No existing corpus combines **multi-decade coverage, multi-outlet breadth, bilingual EN/FR** and **sentence-level supervised annotation** of frames, actors, events and locations.

What this puzzle demands of a research infrastructure

1 EXTRACT EVERY ACTOR

Persons, organizations, types of messengers (scientists, politicians, activists, experts) named or quoted in every sentence.

2 EXTRACT EVERY FRAME

Economic, political, scientific, environmental, health, justice, cultural, security — each with sub-categories for impacts, debate and tone.

3 ANCHOR EVERY CLAIM

Geographic anchor (provinces, cities), events (disasters, conferences, elections), solutions (mitigation, adaptation) and emotional tone.

RESEARCH QUESTION

Can we build a multi-decade, multi-outlet, bilingual climate corpus where every sentence has been parsed for actors, frames, events and locations, at near-human accuracy?

PART II

Method

The corpus · the pipeline · annotation in practice

The CCF Database: Canadian Climate Framing

266K

Articles

1978 – 2024

20

Newspapers

EN + FR

9.2M

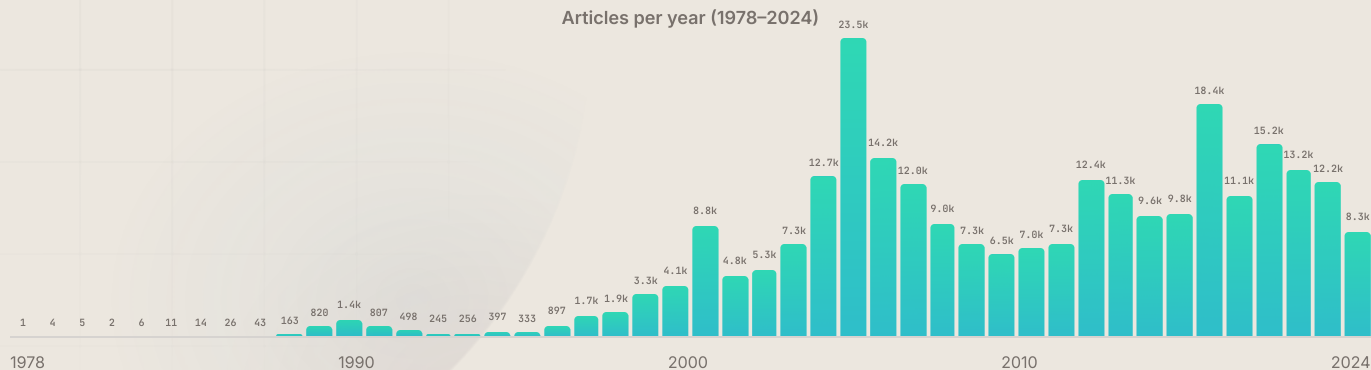
Annotated sentences

BERT / CamemBERT

65

Categories

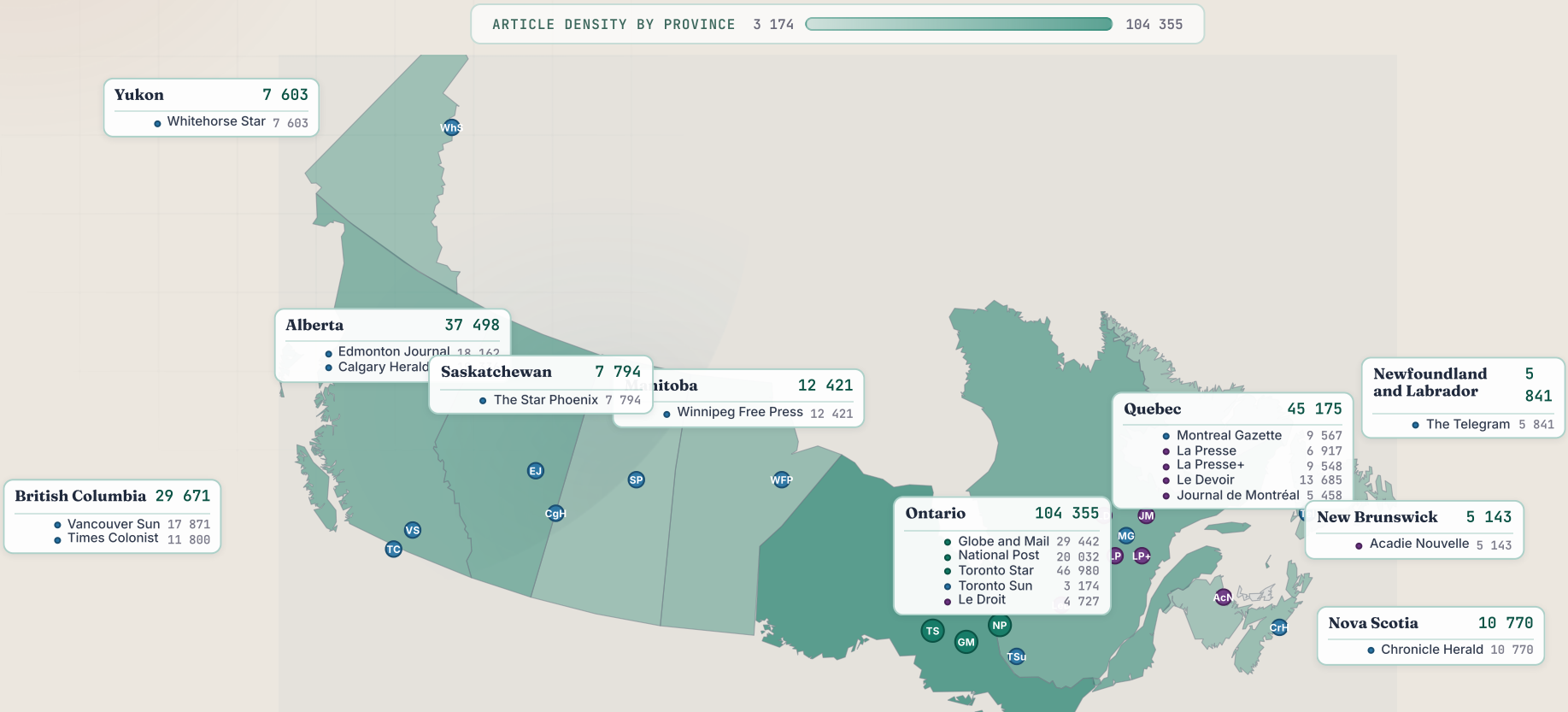
hierarchical



The largest ML-annotated climate discourse corpus in Canada

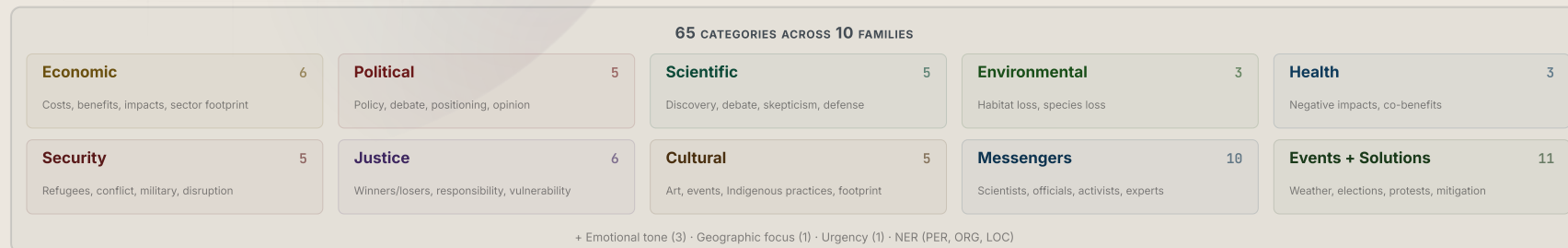
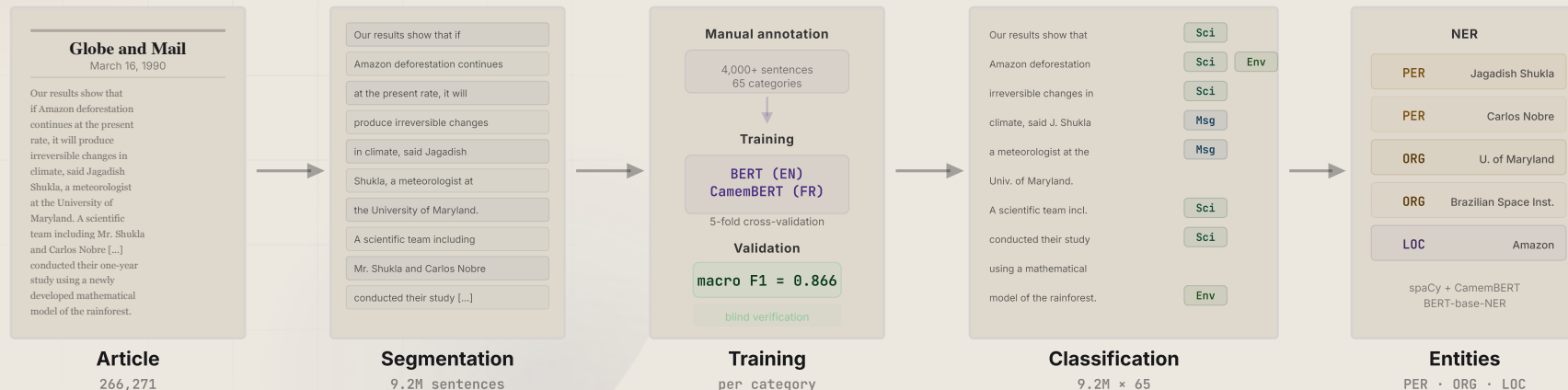
Macro F1 : 0.866 – Lemor, Pillod, Taylor (2025)

All 20 newspapers in the corpus, mapped



Our method in a nutshell

From a single article to 9.2 million annotated sentences across 65 binary categories



65 categories, organised by frame

Annotation in practice

One sentence, four article-level variables: parent frame, two economic sub-categories, and a geographic anchor

Winnipeg Free Press • 21 JULY 2004 • MANITOBA

« By 2050, an average summer temperature in Edmonton could be 1.5 to 4.5 °C higher than present. While Canadians may relish the warmer and longer growing seasons and the lower winter heating costs, the downside will be raging forest fires, insect plagues and heat-stressed livestock. »

LOCATION

Edmonton

Alberta

FRAME

Economic frame

eco_frame

COSTS

Costs of climate action

eco_cost

BENEFITS

Benefits of climate action

eco_benefit

One sentence, four article-level variables: a parent frame and its two economic sub-categories, plus a geographic anchor.

Location

eco_frame

eco_cost

eco_benefit

Three examples of questions we can answer

1

REGIONAL PATTERNS

How do regional differences in fossil fuel dependence shape media coverage of climate change in Canada?

2

FRAME EVOLUTION

How have thematic frames evolved across nearly five decades of Canadian climate coverage, since 1978?

3

MEDIA CASCADES

Can we measure and detect a media cascade, and gauge its effect on framing?

PART III

Applications

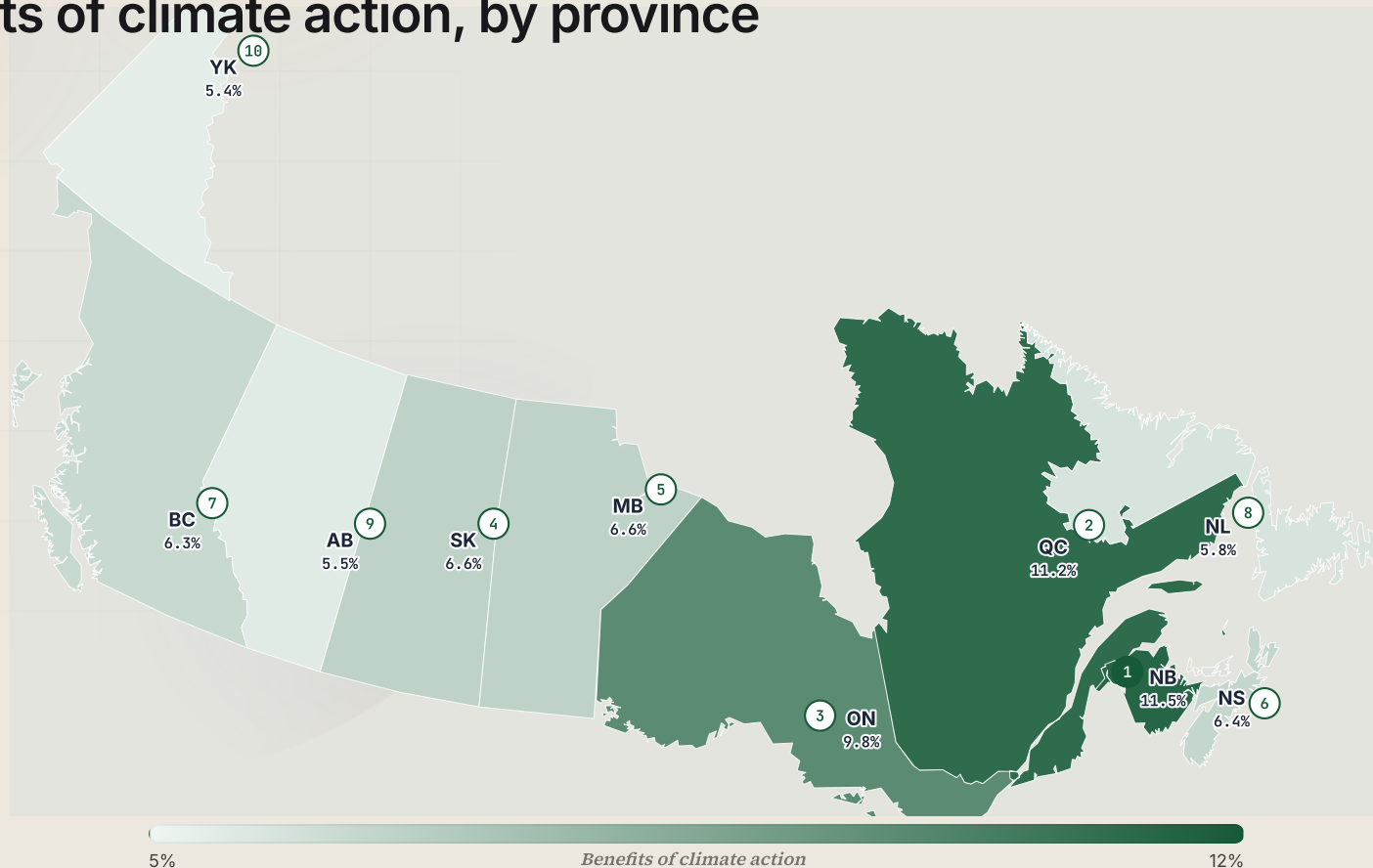
Frame evolution · regional patterns · media cascades

Evolution of thematic frames (1978-2024)

Average sentence share per frame, by year, across the full corpus



Benefits of climate action, by province



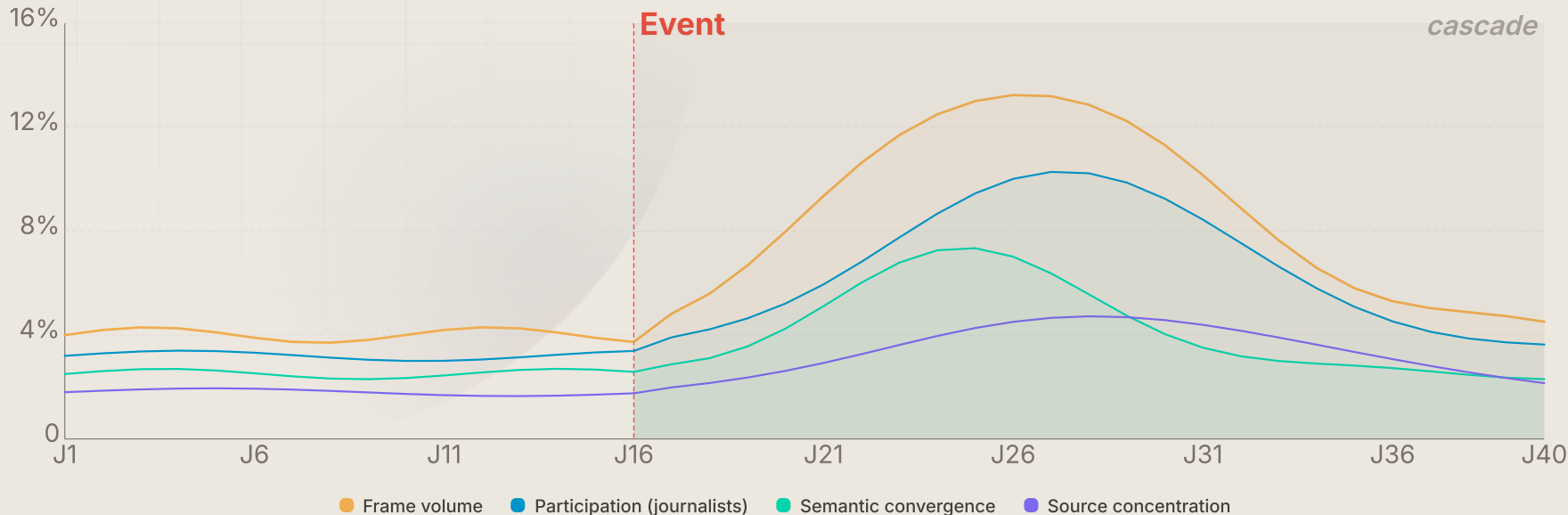
Mean proportion of sentences per article, 2005–2023. Clusters reveal: Quebec → Prairies → Rest of Canada → final ranking.

What is a media cascade?

A frame propagating massively across outlets after a triggering event

Media cascade =

a topic that spreads rapidly across media and journalists, with **convergence** of content, an **increase** in coverage volume, and the arrival of **new actors** in the debate

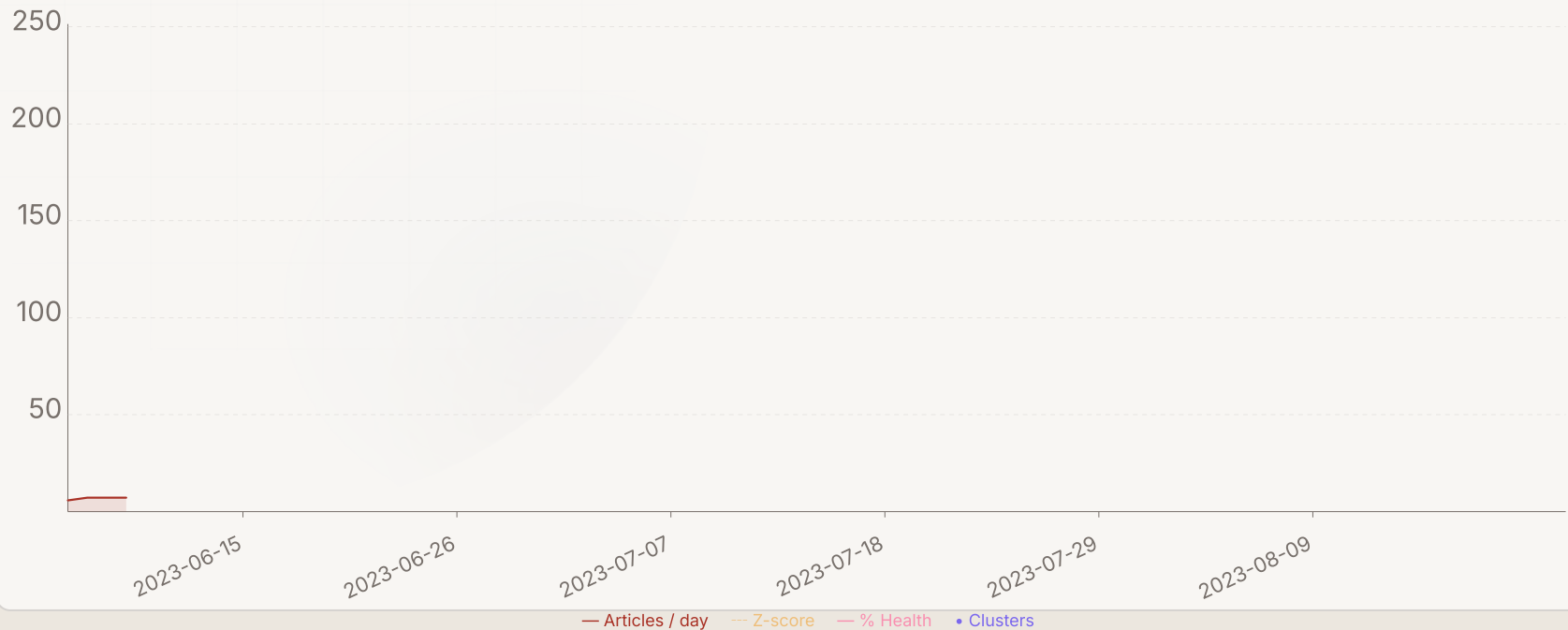


Health cascade, Summer 2023

0.612

MODERATE CASCADE

453 articles 300 journalists 19 outlets 40 days



Three takeaways

1 A FIRST OF ITS KIND

The first openly released machine-learning-annotated, pan-Canadian climate media database, with 47 years of historical depth and bilingual EN/FR coverage.

2 QUANTITATIVE OR QUALITATIVE

The same database supports a wide range of research questions, in quantitative and qualitative methods alike, through a web explorer and an open API.

3 OPEN SCIENCE

Annotations, embeddings and aggregates are openly released on Zenodo, code on GitHub. The data explorer at data.ccf-project.ca is open during EPSS with code EPSS2026.

Thank you!

Questions and discussion

Antoine Lemor · Alizée Pillod · Matthew Taylor

Université de Sherbrooke · Université de Montréal

CAPP · CIRST · RFICS



ccf-project.ca

PROJECT



data.ccf-project.ca

EPSS2026