

# Computational Methods for Social Sciences

Measuring Political Ideas and Frames

Applications on YouTube political content and Canadian climate media

**Antoine Lemor, Ph.D.**

Postdoctoral Researcher

Université de Sherbrooke · RFICS · CIRST · CAPP



# Outline

1

## Why computational?

what these methods enable

2

## Applications and results

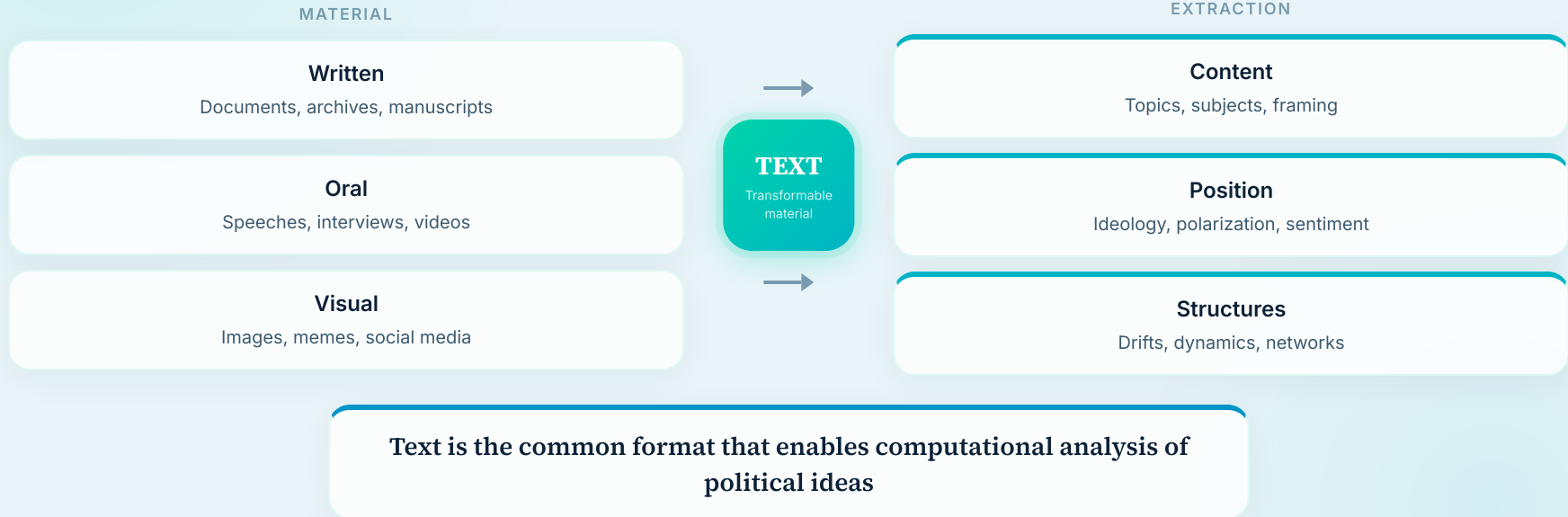
CCF, YouPol, NRx drift

PART I

# Why computational?

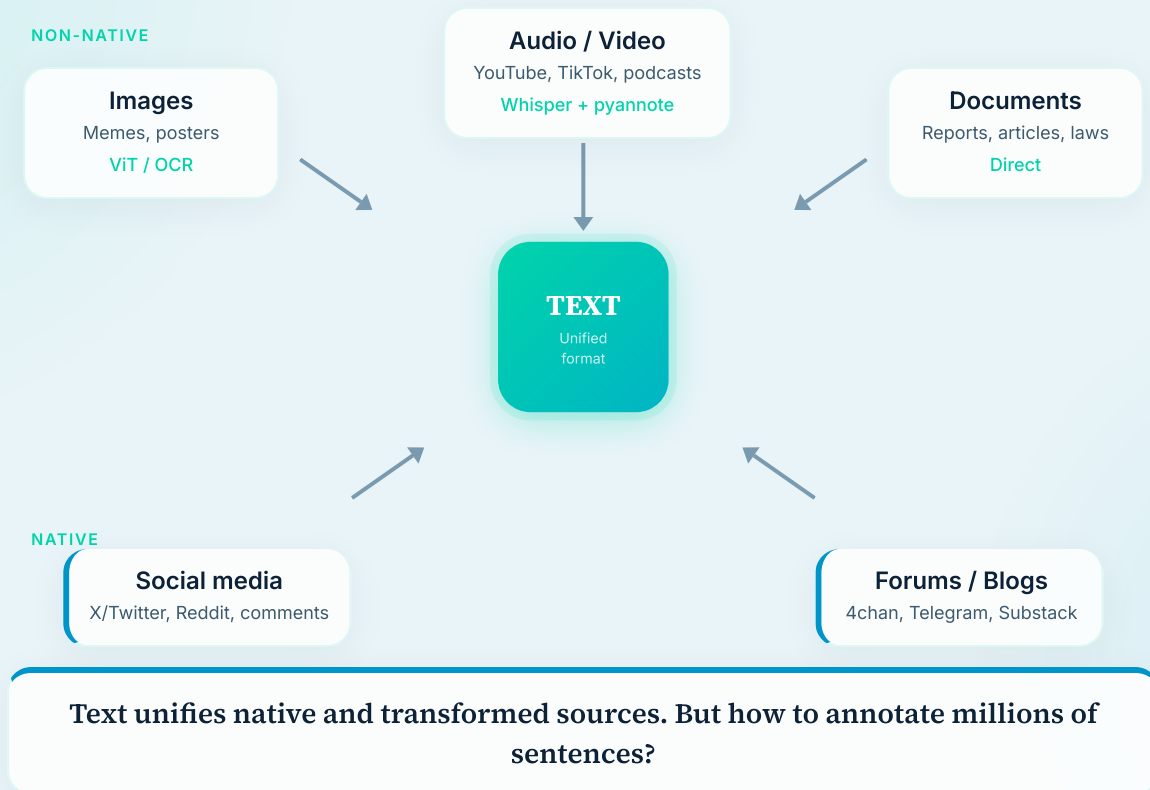
What these methods bring to social science

# Text as foundational material



# From raw content to analyzable text

Text malleability: from diverse sources to analyzable data



# Three operations on ideas

## Classify sentences

*"Nos ancêtres ont bâti cette civilisation par le sang et le sacrifice"*

TRADITION

NATIONALISM

A trained classifier detects ideological motifs or media frames in each sentence, across millions of sentences

## Measure proximity

**0.87** cosine to NRx prototype

Embeddings place each sentence in a vector space. We measure its distance to neo-reactionary ideas, independently of any classifier

## Detect a drift

**+1.13 pp** /decade (panel FE,  $p < 0.001$ )

A panel with fixed effects across 19 channels x 36 quarters isolates the underlying trend. Inaudible, but measurable over 335K sentences

**Computational methods make it possible to measure the evolution and prevalence of specific ideas across millions of sentences**

# LLM Tool: annotating millions of sentences

## The bottleneck

To classify millions of sentences, you need a trained classifier. To train one, you need annotated data. To annotate, you need humans: slow, expensive, not reproducible.

## The solution: LLM Tool

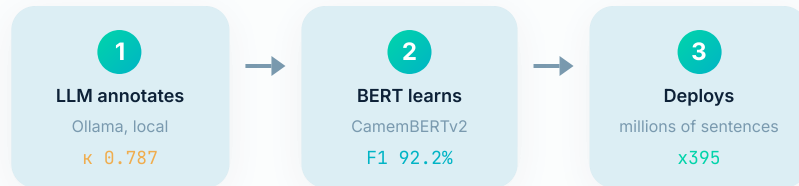
Open-source platform I designed to solve this problem. An LLM annotates a sample following your codebook, these annotations train a fast BERT model, then BERT classifies the full corpus.

Lemor, Dinan, Gilbert (2025) – F1 66.73

## Built for researchers

No code, fully local (sensitive data), 50+ models, 100+ languages, built-in validation (Kappa, F1, Alpha)

## DISTILLATION PIPELINE



## Knowledge distillation

The quality of a large model at the speed of a small one. One LLM annotates 1,000 sentences so BERT can annotate 3 million.

Used for **CCF** (266K articles, 9.2M sentences) and **YouPol** (24.7K videos, 2.95M sentences)



PART II

# Applications

Measuring frames, detecting drifts

# Two databases built with these methods

Throughout this presentation, I will show two concrete applications of the tools and methods just described. Both databases are accessible via the QR codes below.

## CCF

[data.ccf-project.ca](https://data.ccf-project.ca)

Canadian climate media framing  
266K articles · 9.2M sentences · 1978-2024



## YouPol

[data.you-pol.com](https://data.you-pol.com)

Political video content (YouTube, TikTok)  
24.7K videos · 2.95M sentences · FR + QC



TEMPORARY PASSWORD

**IDEOLOGYWORKSHOP**

# Application 1: the CCF database

**266K**

Articles

1978 – 2024

**20**

Newspapers

EN + FR

**9.2M**

Annotated sentences

BERT / CamemBERT

**65**

Categories

hierarchical

Articles per year (1978–2024)



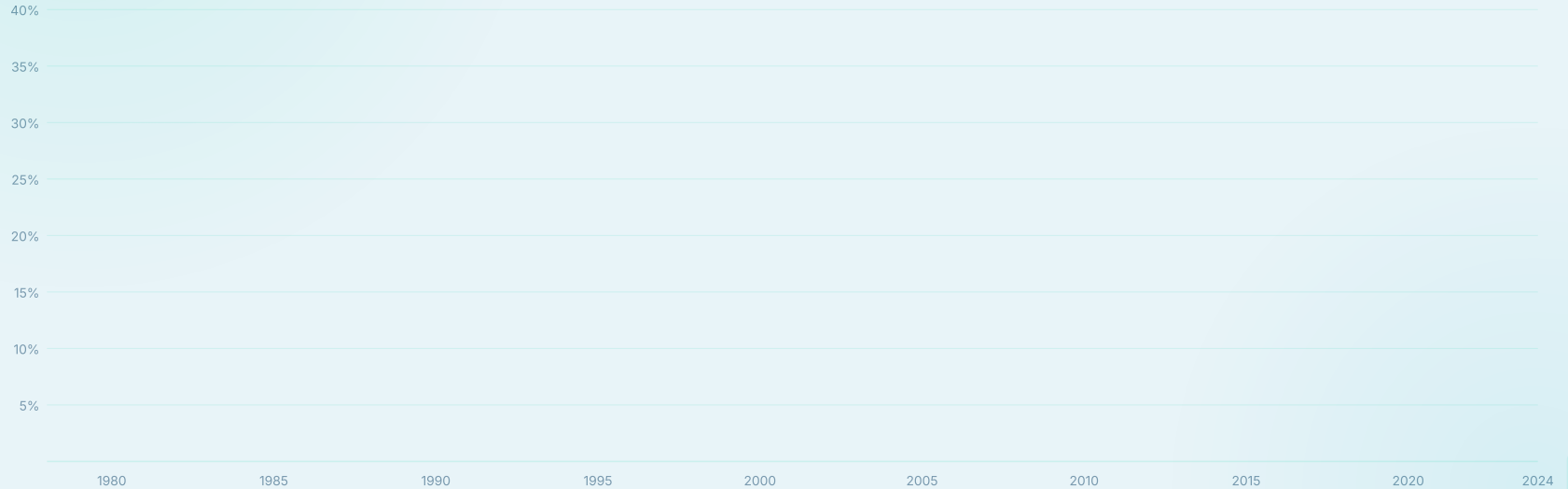
**The largest ML-annotated climate discourse corpus in Canada**

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



# Evolution of thematic frames (1978-2024)

Average proportion of sentences per frame per year



[data.ccf-project.ca](https://data.ccf-project.ca)

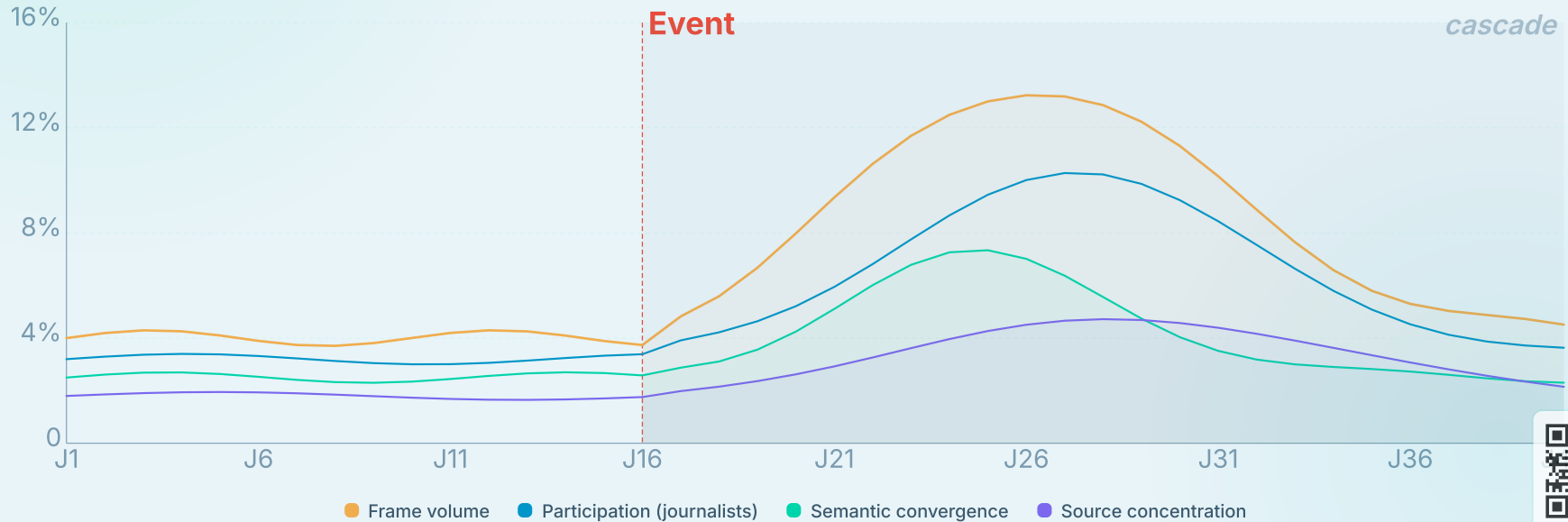
IDEOLOGYWORKSHOP

# What is a media cascade?

When a media frame propagates massively across outlets

## Media cascade =

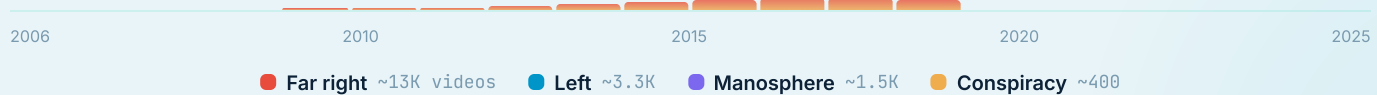
a topic that spreads rapidly across media and journalists, with **convergence** of content, an **increase** in coverage volume, and the arrival of **new actors** in the debate



data.ccf-project.ca

IDEOLOGYWORKSHOP

# Application 2: YouPol



**Collaborative research infrastructure for political video content**

Lemor & Boursier (2026) – you-pol.com – data.you-pol.com



data.you-pol.com

IDEOLOGYWORKSHOP

# From video to data: the pipeline

7 automated steps, from raw video to annotated sentences

## 1. POLITICAL VIDEO (YOUTUBE / TIKTOK)



*"Why mass immigration destroys our civilization"*

Far-right channel · 45 min · 180K views



## 2. AUDIO PROCESSING & TRANSCRIPTION



## 3. DIARIZED SEGMENTS

- L0C-1 *"La France n'a pas vocation à accueillir la misère du monde..."*
- L0C-2 *"...nos ancêtres ont bâti cette civilisation par le sang et le sacrifice..."*
- L0C-3 *"...la technologie nous permet maintenant de restaurer l'ordre naturel."*



## 4. NLP ANNOTATION (LLM TOOL + CAMEMBERT)

**CamemBERTv2** Classifiers trained through knowledge distillation

Political Immigration Tradition Nationalism NRx Ecology Technology Gendered rhetoric Authority



data.you-pol.com

IDEOLOGYWORKSHOP

# What YouPol measures

11 thematic motifs, 3 parallel annotation projects

*A CamemBERTv2 classifier trained via LLM Tool identifies political sentences before thematic annotation*



**11 thematic motifs** 5 NRx + 6 classical

## NRX MOTIFS (NEO-REACTIONARY)

- **Ecology** Reactionary ecology, anti-modernity
- **Equality** Critique of egalitarianism
- **Fictional metaphors** Literary/mythical references
- **Libertarianism** Anti-state, techno-capitalism
- **Technology** Accelerationism, transhumanism

## CLASSICAL MOTIFS

- **Authority**
- **Democracy**
- **Immigration**
- **Nationalism**
- **Progress**
- **Tradition**

## 3 parallel annotation projects

### SIED

Socio-Ideological Score

11 macro-categories of far-right discourse

### GENRE

Gendered rhetoric

4 dimensions of gendered discourse

### NR

Neo-reactionary

Technophilia and reactionary politics

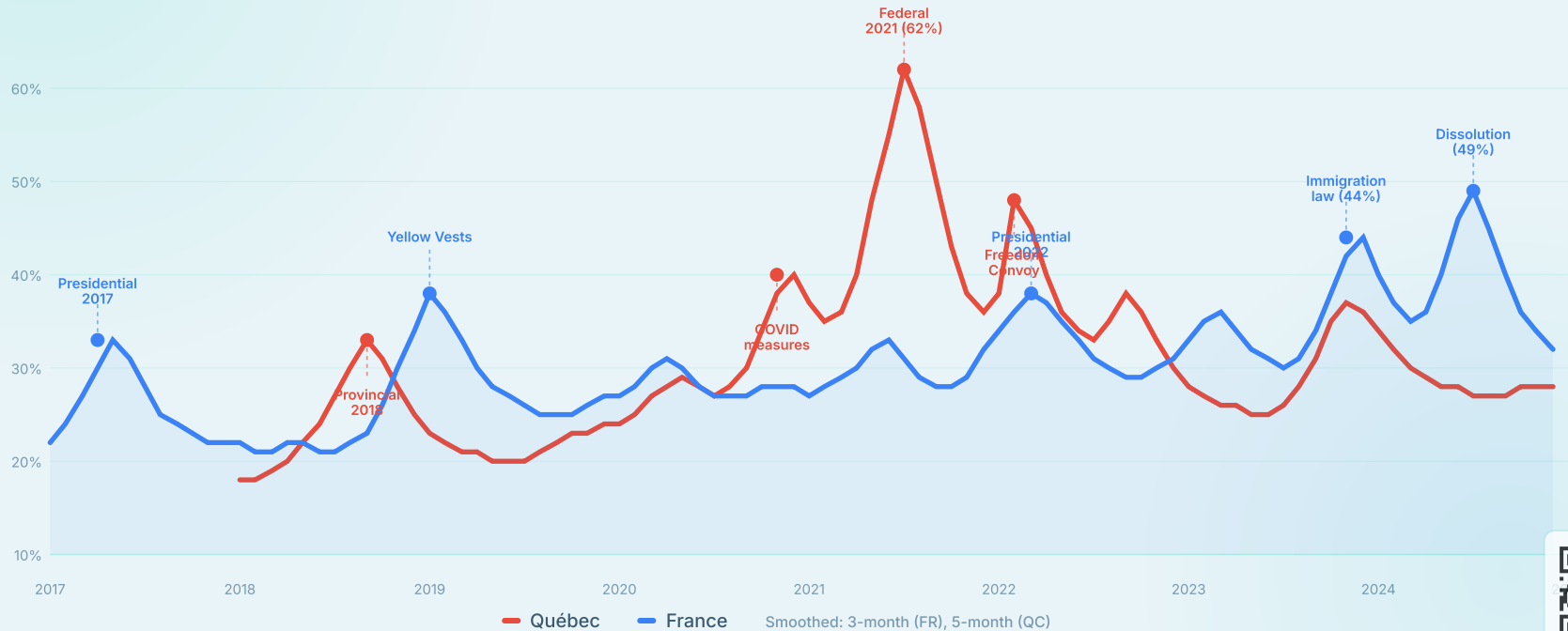


data.you-pol.com

IDEOLOGYWORKSHOP

# Political density over time (2017–2024)

Monthly share of sentences classified as political by detect\_pol — France vs. Québec

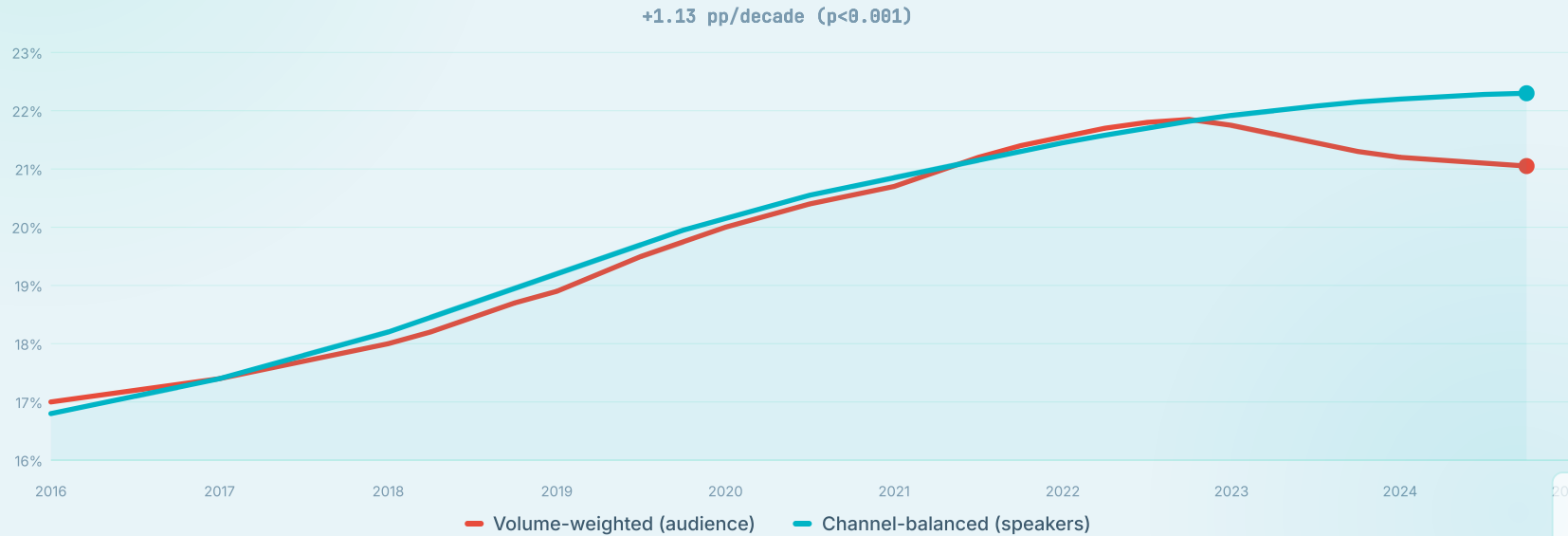


data.you-pol.com

IDEOLOGYWORKSHOP

# What YouPol reveals: NRx drift (2016–2025)

Average NRx motif score across 335K political sentences, 19 stable far-right channels

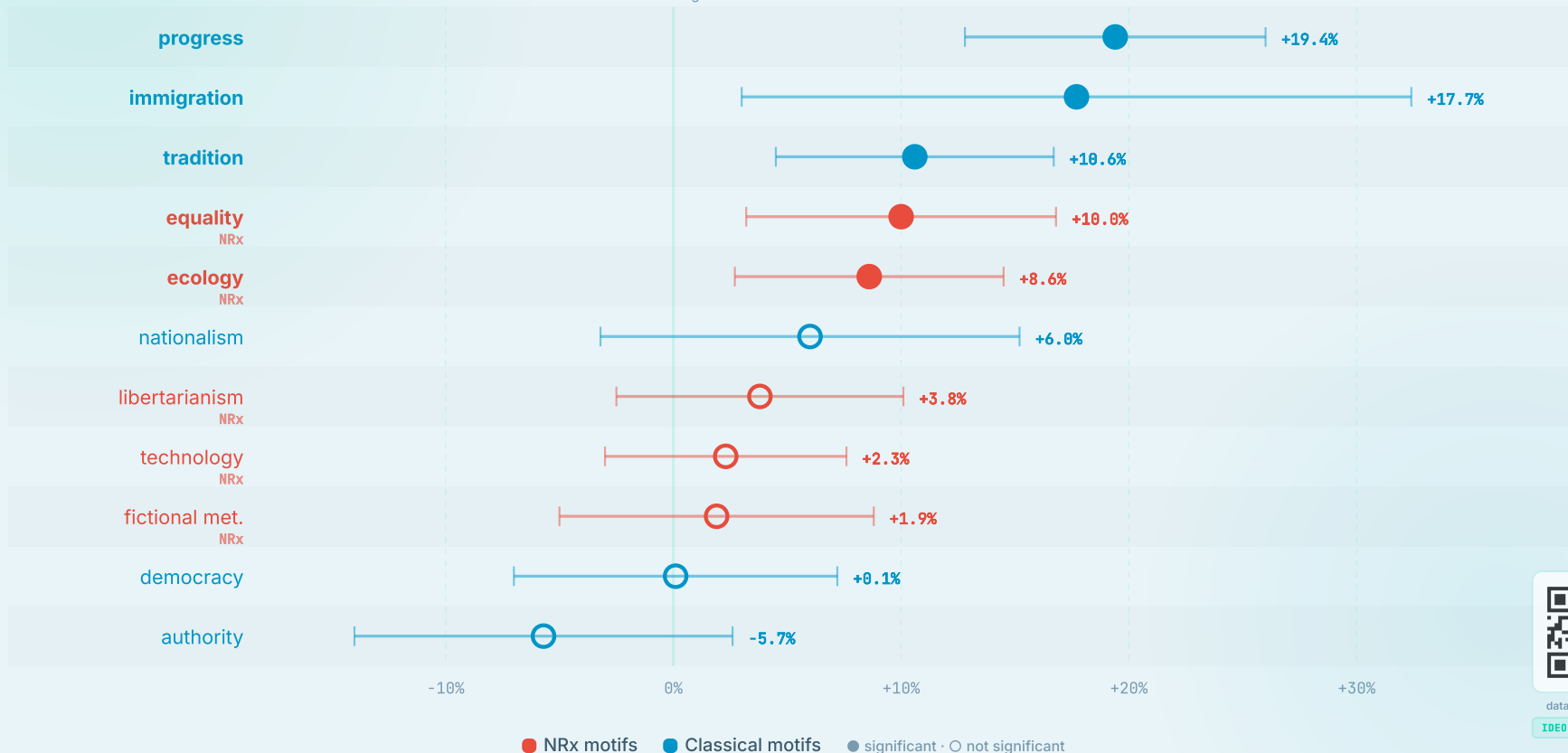


data.you-pol.com

IDEOLOGYWORKSHOP

# All themes and political engagement

Effect of each motif on comment politicization (% change per +1 SD, PPML with channel FE)  
no change



data.you-pol.com

IDEOLOGYWORKSHOP

# YouPol: infrastructure and access

1

## Preserved content

**2 305**

deleted videos

**1.58M**

comments

Far-right: 11.1% removed  
Left: 1.6%

2

## Longitudinal tracking

Full engagement history over time,  
not a single snapshot

3

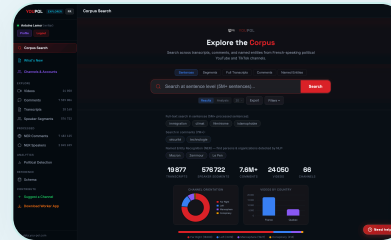
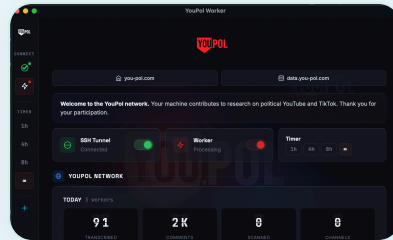
## YCCN

Distributed transcription network.  
Researchers contribute their  
machines

4

## Open API

[data.you-pol.com](https://data.you-pol.com)  
REST API, search, export



[data.you-pol.com](https://data.you-pol.com)

IDEOLOGYWORKSHOP

# Thank you!

Questions and discussion

---

Antoine Lemor, Ph.D.

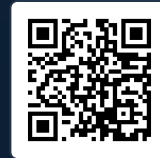
Université de Sherbrooke · RFICS · CIRST · CAPP



[antoinelemor.com](https://antoinelemor.com)



[you-pol.com](https://you-pol.com)



[LLM Tool \(GitHub\)](#)