Automatic Classification of News Subjects in Broadcast News: *Application to a Gender Bias Representation Analysis*

Valentin Pelloin[1] Lena Dodson[2] Émile Chapuis[1]
Nicolas Hervé[1] David Doukhan[1]

[1]INA (Instiut national de l'audiovisuel), France [2]ARCOM (Autorité de régulation de la communication audiovisuelle et numérique), France

Introduction

- Are women equally represented along all subjects (politics, weather, sports, ...) in French broadcast news?
- **Objectives:**
  1. Classification of unsegmented audiovisual broadcast news into topic subjects.
  2. Estimating gender representation biases in French audiovisual media.
- **Constraints:**
  1. Low inference cost to compute analysis on a large scale.
  2. Low human annotation budget for creating ground truth.
- Related works:
  - **GMMP** reports (36 hours): highly detailed (unseen characters, job, importance)
  - **ARCOM** reports (31.8k hours): less detailed, but done on large quantity of data

Data description

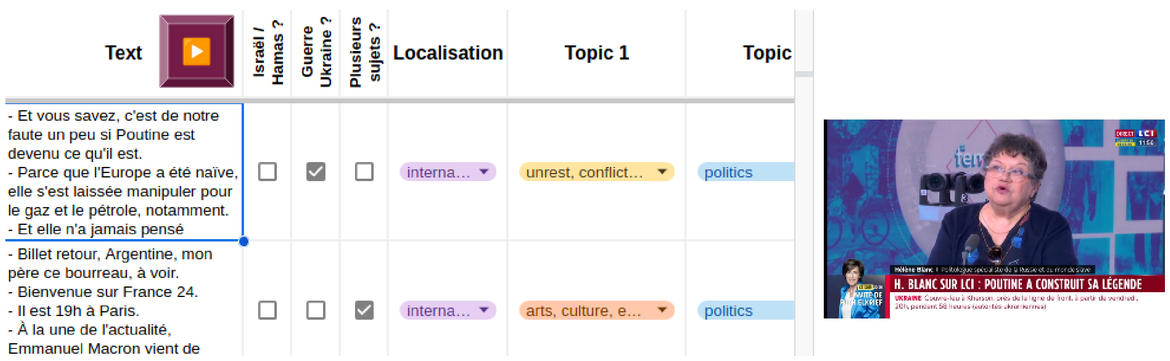
- In France, **ARCOM** is in charge of collecting **content reports from channels** to measure women representation on TV and Radio.
- **In 2023** (May and October) **41 channels** reported **29,707 programs** with type (Information/News, Entertainment, ...) and the number of men/women present in each program.
- We select **11.7k hours of audiovisual data** under either category "*Information/News*" or from 24/7 news channels.
- We **transcribe** the news with *whisper-large-v3* and **merge sequences into dialogues** (max. dialogue length of 60s, max. gap between two sequences of 10s).

Annotation campaign

- We annotate **804 dialogues (3h44min)**, separated into Dev (54min) and Test (2h50min): four annotators, each dialogue is annotated by two people.
- **Multi-label topic annotation** using 18 categories inspired by the IPTC (International Press Telecommunications Council):
  - ↪ unrest, conflicts, war
  - ↪ politics
  - ↪ commercial
  - ↪ weather
  - ↪ ...
- **Localisation** – We indicate the reach of the news subject: *Local*, *National*, *European*, *International*, inspired by the GMMP.
- **Multiple subjects** – We indicate if the dialogue contains multiple unrelated subjects.
- **Wars** – We indicate if the dialogue is about two main mediatized subjects in 2023: *Ukraine-Russia* war or the *Israel-Hamas* war.
- The **global inter-annotation agreement** (Krippendorff's alpha) is at **0.60**, with high agreements for classes such as *commercial*, *unrest/conflicts/war* or *weather*, and low agreements for classes like *social issue* or *science/technology*.

✨ We release the dataset free of charge for research purpose! ✨

Annotation interface



Methodology

- Three classifier model for the topic annotations:
  1. **Baseline**: BERT token-classification
  2. **LLM**: Mixtral-8x7B prompted classification
  3. **Teacher/student model**: BERT token-classification finetuned on 353k dialogues annotated by Mixtral-8x7B (2.)
- For the whole 11.7k hours / 2.1M dialogues dataset, we:
  - use inaSpeechSegmenter to automatically detect gender speaking times
  - use the best model (Teacher/Student - *CamB*) to perform topic-classification

Results: automatic system evaluation

	Model	Micro (%)			Macro (%)		
		F1	P	R	F1	P	R
Baseline	CamB	50.5	77.2	37.5	24.3	82.4	20.8
	CamL	58.5	73.0	48.8	37.3	82.5	31.8
	FlauLC	55.6	69.0	46.5	40.0	69.6	34.2
Mixtral-8x7B		58.6	63.0	54.8	53.8	59.8	51.5
Teach./Student	CamB	**62.5**	71.7	55.3	**58.5**	68.9	53.0
	CamL	60.9	71.7	53.0	55.9	67.3	49.8
	FlauOF	61.3	69.0	55.1	56.5	64.3	52.6
	FlauOM	62.0	74.5	53.2	54.8	67.3	48.4
	FlauOA	**62.6**	73.3	54.5	55.9	68.1	49.3

Table 1. F1-score, Precision (P) and Recall (R) on the annotated test set.

👉 **Mixtral LLM** obtains **better results** than **Baseline BERT**

👉 Teacher/Student: **Student outperforms the Teacher**, lower inference cost (120.7x faster on 9.6x cheaper hardware)

Results: gender bias analysis

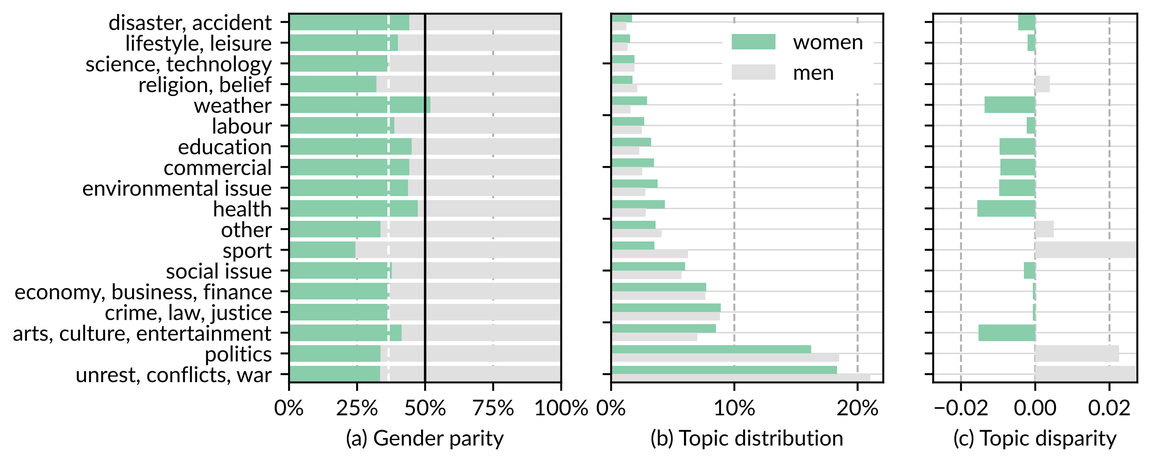


Figure 1. Measured gender representation bias per topic

- Globally, women have a speaking time of 36.58% compared to men.
- [Fig 1.a] We notice the speaking times for **women** is :
  - 👉 **lower than average** for ***sport*** (24.5%), *religion/belief* (32.2%), ...
  - 👉 **higher** for ***weather*** (52.0%), *health* (47.4%), *education* (45.1%), ...
- [Fig 1.b/c] At equal speaking time :
  - 👉 **men** are more **likely** to speak of **armed issues** (*unrest/conflicts/war*)
  - 👉 **women** are more **likely** to speak of ***arts/culture/entertainment***

Limitations

- Difficult annotation: 0.60 global inter-annotator agreement. The model used for the gendered bias analysis gets 62.5% micro-F1.
- Dataset and study done on only two months of 2023.
- Strategic essentialization: gender is assumed to be binary and is determined automatically without inquiring how individuals identify.

Code & dataset: Paper:

Interspeech 2024 - Kos, Greece

Contact: vpelloin@ina.fr
Paper id: 1854