
Les Françaises et Français face à l'information

Max Chateau - Rasim Erben - Myndie Ferrandez - Quentin Jacquot

Polytech Montpellier - DaMS4



1 Méthodologie -

1.1 Enjeux et objectifs du projet -

Récemment, un article faisant la une du 13 Janvier du *Monde* titrait *Le Rassemblement national et ses idées s'installent dans l'opinion*. Toutefois comme le pointent des sociologues comme Félicien Faury ou Vincent Tiberj, les médias, en particulier la télévision, jouent un rôle essentiel dans la diffusion d'idée dans la société. En parallèle, plusieurs études et rapport montre un décalage entre la société française et ses médias, on peut évoquer le temps de parole des femmes qui en 2023 ne représentait uniquement 34% à la télé et la radio.

De surcroît, le paysage audiovisuel français connaît de nombreuses transformations. La télévision et la radio, longtemps considérées comme les principales sources d'information, sont aujourd'hui concurrencées par les réseaux sociaux et les plateformes numériques. Une part importante des Françaises et des Français s'informe désormais en ligne, ce qui modifie les habitudes de consommation de l'information et peut influencer la confiance accordée aux médias traditionnels.

Ces constats soulèvent des questions importantes sur les biais présents dans l'offre d'information et sur leurs effets possibles sur le rapport des citoyens à l'actualité. Dans ce contexte, ce projet a pour objectif d'analyser les transformations du paysage audiovisuel français, en s'intéressant à la fois aux contenus diffusés (thèmes traités et représentation des femmes) et au rapport des Français à l'information (pratiques et confiance). La problématique centrale de ce travail est donc la suivante : dans quelle mesure les transformations du paysage audiovisuel français influencent-elles le rapport des Français à l'information, notamment en termes de pratiques et de confiance ?

Nous traiterons ainsi dans un premier temps de l'analyse thématique concernant les chaînes de TV, puis nous verrons

1.2 Les Datasets utilisés -

Les données du challenge

Sur le site du challenge, nous avons exploité les datasets suivants :

- ♦ Audience de la télévision - CNC
- ♦ Consommation VOD - CNC
- ♦ Télévision de rattrapage - CNC
- ♦ Classement thématique des journaux TV - INA

Par manque de temps nous n'avons pas pu analyser en détail le jeu de données de l'ARCOM et en extraire le contenu concernant *Les français et l'information*, qui avait l'air de contenir des données intéressantes pour corroborer notre analyse, mais n'étaient pas aussi simple à exploiter que les autres jeux de données.

Le site data de l'INA

L'INA met à disposition de tout le monde le site data.ina.fr, ce dernier regroupe les analyse d'une IA sur les flux audios des principales chaines de la télévision. Les chaines d'information en continue présentes sont : BFM TV, CNews (anciennement I-Télé), France Info, et LCI. Et celles des chaines généralistes sont : TF1, France 2, France 3, M6 et Arte.

Différents axes sont présents pour chaque chaines de manière journalière :

- ◆ Proportion temps de parole Femmes/Hommes
- ◆ Nombre de fois qu'un mot a été prononcé
- ◆ Top 50 des personnes évoquées

Nota Bene : Ces données ne sont (en théorie) pas accessibles de manière brut publiquement, on a donc développé un petit programme pour communiquer avec l'API de l'INA et ainsi récupérer les données (dispo sur Github).

Jeux de données annexes ou non exploités

On a contacté Médiamétrie afin d'essayer d'obtenir des données sur les audiences plus précises, sans réponse de leur part malheureusement...

On s'est également basé sur l'actualité afin de pouvoir expliquer certaines données obtenues. On a également utilisé certains sondages lorsque cela nous semblait pertinent pour appuyer notre argumentaire.

2 Analyse thématique -

2.1 Chaines d'information en continue

Dataset utilisé : data.ina.fr

Les données n'étant pas classé par thème initialement, on a établi une liste d'une centaine de termes auxquels on a assigné un thème (ex : on a attribué le terme "Immigration" est associé au thème extrême droite). Cette liste est disponible dans la partie méthodologie de la section thématiques du site rendu. Les limites de cette méthode étant que les listes ont pu être mal établies d'une part (biais idéologique, nombre de termes insuffisants, etc...), d'autre part on n'a pas le contexte dans lequel le mot a été prononcé.

Evolution des thèmes diffusés

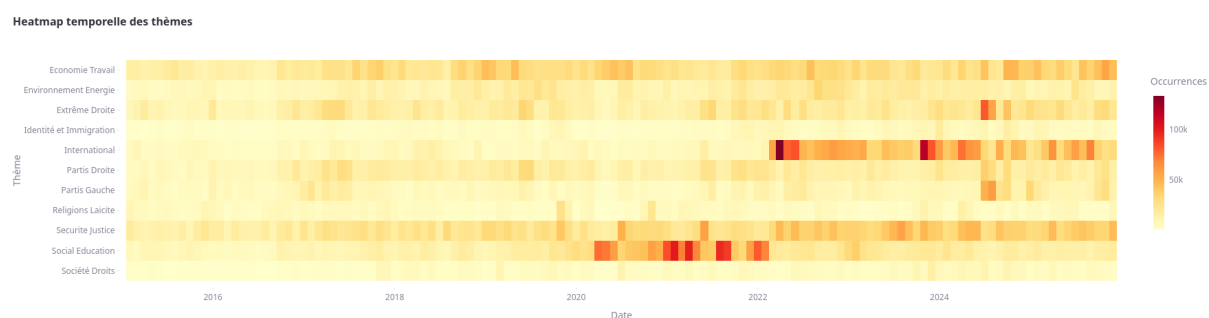


FIGURE 1 – Heatmap temporelle des thèmes sur les chaines d'info en continu

Nota Bene : Un thème "Politique" est présent également, mais les termes présents dedans sont tellement mentionné qu'ils éclipsent les autres thèmes, donc pour cette première analyse on fait fi de ce thème.

La vue globale nous montre plusieurs info intéressantes. Tout d'abord à partir de 2020 on constate une surreprésentation du thème social, notamment dû au fait que les termes liés au COVID sont compris dedans.

On constate également qu'à partir de Février 2022, le thème social est quasiment entièrement remplacé par les thèmes Sécurité/Défense et International, date du début de l'invasion Russe en Ukraine.

Un pic réapparaît pour ces mêmes thèmes en octobre 2023, date de l'attaque du Hamas en Israël, puis des attaques successives contre la Bande de Gaza.

Le dernier pic notable est en Juin 2024 concernant les thèmes liés à l'extrême droite, ainsi que la gauche dans une moindre mesure (on reviendra sur ce point plus tard), plus précisément lors du résultat des européennes et après l'annonce de la dissolution de l'assemblée nationale.

A l'inverse on constate que certains thèmes sont moins représentés que d'autres sur le temps long et n'ont que quelques représentation épisodique. On peut prendre l'exemple des sujets liés aux sujets société/droits et l'environnement. Bien que ces sujets semblent être important pour les français, comme le montre le Baromètre de l'IPSOS de septembre 2025. Par exemple, le sujet de l'environnement apparait comme sous représenté, or il apparait comme étant une préoccupation importante pour 20% des français en septembre 2025. Ceci peut être expliqué en partie par le fait que les chaines d'infos sont possédées par des entreprises privées, qui peuvent avoir des intérêts économiques contraires à la mise en avant de ces sujets.

Corrélation entre thèmes et chaines -

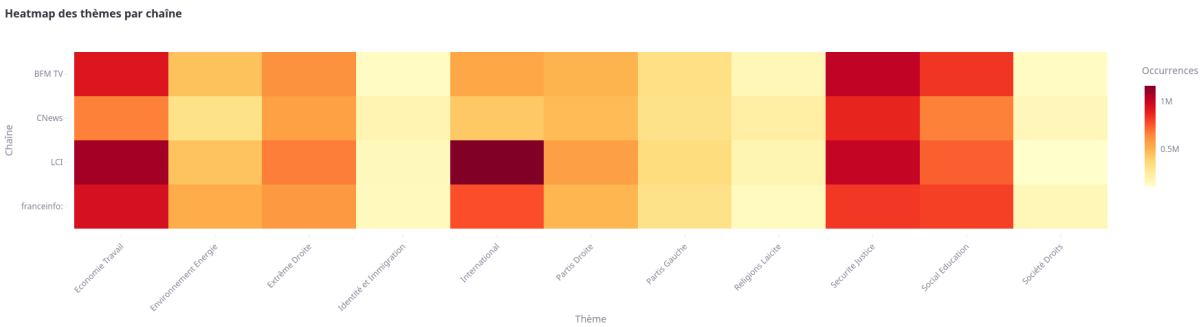


FIGURE 2 – Heatmap de la présence des thèmes sur les chaines d'info en continu, période 2015-2025

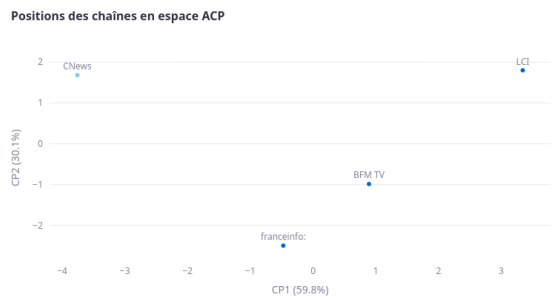


FIGURE 3 – ACP des chaines d'information en continue sur les thèmes, période 2015-2025

apparaître en 2018 (date où Vincent Bolloré reprend en main la chaîne), se calme en 2020 avec le COVID, puis repart en 2021. Tandis que dans le même temps les thèmes liés à l'environnement et aux droits/société sont moins représentés sur la chaîne.

On constate également que la Droite et l'extrême droite sont plus représentés que la Gauche sur l'ensemble des chaînes.

Top 15 thèmes discriminants

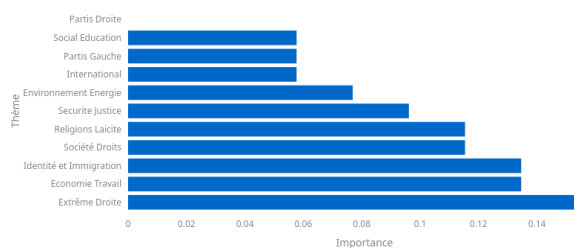


FIGURE 4 – Importance des thèmes pour la classification à l’aide d’une random forest sur toute la période

sur CNews.

On constate lors des échéances électorales que les thèmes politiques prennent une place plus importante sur les chaînes d’infos. Avec une très forte hausse lors des législatives de 2024, notamment au niveau de l’extrême droite notamment.

On ainsi peut re évoquer l’article cité en introduction, qui montrait la progression des idées du Rassemblement National dans l’opinion publique. On peut légitimement se demander si cette progression n’est pas en partie due à la surreprésentation de ces thèmes dans les médias audiovisuels, et notamment