



HAL
open science

Détecter et contextualiser le faux dans les énoncés politiques en télévision : conception d'un dispositif multimodal de vérification

Frédéric Rayar, Jérémie Nicey

► To cite this version:

Frédéric Rayar, Jérémie Nicey. Détecter et contextualiser le faux dans les énoncés politiques en télévision : conception d'un dispositif multimodal de vérification. Infox sur Seine 2023, Mar 2023, Paris, France. hal-04338277

HAL Id: hal-04338277

<https://hal.science/hal-04338277>

Submitted on 12 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détecter et contextualiser le faux dans les énoncés politiques en télévision : conception d'un dispositif multimodal de vérification

Frédéric RAYAR ⁽¹⁾ et Jérémie NICEY ⁽²⁾

⁽¹⁾ LIFAT, EA 6300, Université de Tours, frederic.rayar@univ-tours.fr

⁽²⁾ PRIM, EA 7503, Université de Tours, jeremie.nicey@univ-tours.fr

Les fausses nouvelles, rumeurs et contre-vérités, tout en n'étant évidemment pas nouvelles, constituent un défi contemporain de société par leur prolifération d'une ampleur, elle, inégalée et reposant en grande partie sur les dispositifs numériques. Elles engendrent le brouillage des repères factuels, historiques et sociaux dans l'espace public, à plusieurs niveaux : international, national et local. Les fausses informations, souvent regroupées improprement sous la vulgate de *fake news* (terminologie pourtant problématique à plusieurs égards) modifient et remodelent considérablement le rapport de chacun aux actualités et aux connaissances, rendant ces dernières fragiles, friables et contestées.

Face à ce phénomène, un des enjeux est donc de fournir des repères fiables – mais aussi fonctionnels et rapides – aux consommatrices et consommateurs d'information d'actualité qui, parmi plusieurs contraintes, manquent de temps pour vérifier par eux-mêmes la validité d'un énoncé politique ou d'une rumeur. C'est ce que proposent les productions des unités journalistiques de *fact-checking*, depuis le début des années 2000 aux Etats-Unis et depuis la fin des années 2000 en France. Pour autant, ces dernières sont relativement sous-consommées : seulement par 47% des usagers [Kantar, 2022], et les publics qui y ont recours n'y sont, à l'évidence, pas connectés en permanence.

Enjeux de la détection de fausses informations télévisées

La détection des fausses informations est ainsi devenue cruciale mais elle présente, dans ses configurations actuelles, plusieurs limites. D'une part, face à la volumétrie/ampleur des contenus à vérifier, elle est souvent effectuée plus ou moins manuellement par les *fact-checkers*, et n'est évidemment pas exhaustive, ni suffisamment réactive. D'autre part, quand elle n'est pas manuelle et qu'elle profite des dispositifs des réseaux sociaux, la détection ne bénéficie pas d'efforts optimaux de ces derniers (qu'il s'agisse de l'outil CrowdTangle, par exemple désactivé en 2022 [Smalley, 2022] ; qu'il s'agisse de l'exclusion délibérée de la thématique politique par certains acteurs, particulièrement Facebook ; ou encore qu'il s'agisse des opérations, certes intéressantes mais imparfaites et non transparentes, pour signaler rétroactivement à un utilisateur de Twitter ou Facebook qu'il a consulté un contenu biaisé). Enfin, cette détection est le plus souvent limitée aux contenus numériques *textuels* (articles en ligne, extraits de sites, de blogs et de *posts* sur les réseaux sociaux). De fait, les contenus d'information *vidéo* (journaux télévisés, émissions, interviews en direct, etc.), dont la consommation dite « linéaire » (i.e. au moment de leur diffusion) reste prégnante, ne donnent pas lieu, pour l'heure, à une détection satisfaisante.

Nous proposons ici d'approcher ces **contenus informatifs audiovisuels**, et de tenter de les confronter à un **dispositif se donnant pour objectif d'automatiser, dans une certaine mesure, l'identification et la contextualisation des fausses informations** qu'ils contiennent potentiellement. Nous fondons le principe de cette identification sur la logique suivante : **l'antériorité des énoncés politiques faux, biaisés ou douteux**. Étant donné que ces énoncés abusifs se révèlent être récurrents et répétés, et étant donné que les thèmes sensibles et propices à de tels énoncés (ex : immigration, chômage, controverses environnementales, polémiques judiciaires, etc.) sont eux-mêmes abordés de manière récurrente dans ces émissions audiovisuelles, il est envisageable d'utiliser l'antériorité (à un temps T moins 1 an, T moins 2 mois, etc.) de formulation des rumeurs, mensonges ou approximations par des personnalités politiques, pour **tenter de développer un cadre – semi-automatisé – d'alerte**, à un temps T, à destination des

consommateurs de l'actualité télévisée, **mais aussi de contextualisation/explication, et de prévention.**

Approche

Nous proposons de présenter dans cette communication scientifique les premières tentatives et hypothèses de conception d'un tel outil de vérification, semi-automatique et multimodale, d'énoncés politiques dans des vidéos. La Figure 1 ci-dessous présente de manière simplifiée notre approche.

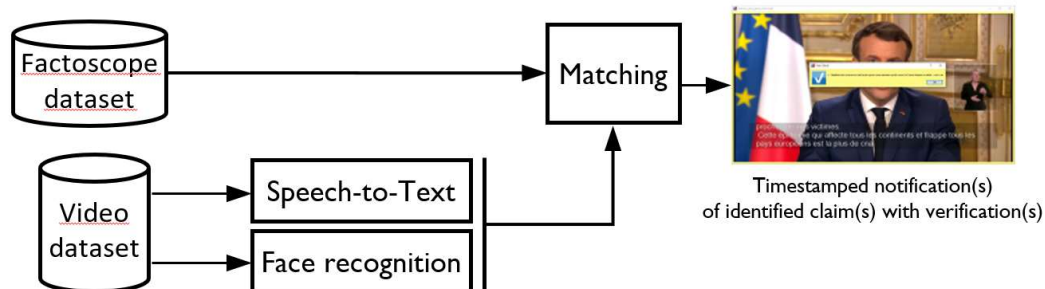


Figure 1. Workflow de l'approche proposée.

Nous soulignerons les opportunités qu'une telle conception fait apparaître mais aussi les contraintes auquel elle doit faire face, notamment :

- le traitement de vidéos (en diffusion télévisée dite « linéaire » ou en *replay*) ;
- le choix de ne pas procéder à une détection via *worthwhile checking claim* ;
- la mise en place d'un *matching* avec une base existante d'énoncés politiques (affirmations sensibles, contre-vérités, approximations, éléments de langage, etc.) ;
- la limitation à la langue française et au contexte français d'actualité politique qui, pour l'heure, n'ont d'ailleurs pas fait l'objet de travaux de recherche sur la détection.

Base de données

Dans un premier temps, une base existante d'énoncés vérifiés est utilisée : FactoScope : nous nous appuyons sur « FactoScope 2017-2022 »¹, une initiative en partenariat avec LeMonde.fr. Il s'agit d'un blog et site d'agrégation livrant des articles de *fact-checking* s'intéressant particulièrement au monde politique, et réalisés ou compilés par les journalistes-étudiants de l'École publique de journalisme de Tours (EPJT). Ce site étant construit avec un système de gestion de contenu web (nommément WordPress), nous avons créé un *web scraper* permettant de récupérer et de structurer ces articles au sein d'une base de données.

Ainsi, nous avons récupéré plus de 1200 faits ou énoncés vérifiés, traitant de sujets répartis sur 12 thèmes variés (éducation, environnement, justice, santé, sécurité, etc.), et qui ont été énoncés par 200 personnes politiques différentes durant les 5 dernières années (2017-2022). Trois niveaux de véracité ont été considérés et la base est composée de 55% de faits erronés, 25% de faits imprécis et 20% de faits corrects. Le travail de vérification est sourcé et renvoie vers des liens de services de *fact-checking* nationaux de renom tels que les unités AFP-Factuel (Agence France Presse), Les Décodeurs (Le Monde), CheckNews (Libération) ou encore Le Vrai du faux (France Info). À partir du contenu textuel, si nous portons notre attention par exemple sur le seul attribut 'description' de la base FactoScope récupérée, nous disposons d'un corpus de 241 202 mots (dont 17 998 uniques), contenant 23 886 entités nommées (principalement des noms de personnalités, d'organisations, de

¹ <https://rattrapages-actu.epjt.fr/factoscope>

postes, de lieux et des dates). La base FactoScope est mise à disposition de manière ouverte et structurée au format xml².

Dans un second temps, nous avons constitué un corpus multimédia (vidéo/audio) d'émissions politiques grâce à une campagne d'acquisition menée durant 3 mois (printemps 2022) sur 8 chaînes distinctes de la TNT. Après raffinage de cette base, nous disposons de $\approx 6\,730$ émissions télévisuelles représentant 6 540 heures de vidéo pour environ 2To. Ce contenu multimédia est aussi mis à disposition des chercheurs³ et plus de d'informations sont données dans [Rayar et al, 2022]. La base multimédia n'ayant pas été disponible lors de la réalisation des premiers travaux, nous présentons ici les résultats sur un jeu de données synthétique et donc en mode *offline* : il s'agit d'une quinzaine de capsules vidéo récupérées et où des faits énoncés, quelle que soit leur véracité, apparaissent dans la base FactoScope.

Prototype

À partir des vidéos, deux types d'éléments d'intérêt sont extraits afin de pouvoir réaliser une recherche multimodale dans la base FactoScope : les visages des personnalités politiques et les transcriptions. Concernant le module de reconnaissance faciale, nous avons utilisé FaceNet [Schroff et al., 2015], un algorithme à base de réseaux de neurones proposé en 2015 et présentant des taux de performance supérieurs à 95%, pouvant aller jusque 99%, sur des jeux de données de référence. Afin de pouvoir l'utiliser dans notre cas, une étape de *fine-tuning* a été réalisée afin de reconnaître les 10 personnalités politiques les plus présentes dans la base FactoScope. Concernant les transcriptions, celles-ci ont été obtenues à l'aide d'un outil de reconnaissance automatique de la parole (*speech-to-text*) libre, nommément Vosk⁴ et en utilisant leur modèle léger pré-entraîné sur la langue française. La qualité des transcriptions obtenues reste relativement correcte pour être utilisées dans notre *workflow*. Outre les transcriptions textuelles, un horodatage (*timestamp*) de chacun des mots reconnus est récupéré afin de pouvoir être utilisé pour afficher des notifications.

Fort de la disposition de ces 2 types d'éléments d'intérêt, un algorithme en 2 étapes a donc été implémenté : la base de faits/sujets est filtrée en ne sélectionnant que ceux qui ont été évoqués par la personnalité politique reconnue, puis une technique de traitement automatique des langues, nommément le *word embedding*, a été utilisée pour mettre en correspondance un fait avec la base filtrée. Ainsi, disposant désormais de l'identité de la personnalité publique, de la date où l'affirmation a été prononcée et du *fact-check* de cet énoncé, nous avons ainsi la possibilité de faire apparaître un *pop-up* contenant un certain nombre d'informations visuelles lors de la lecture vidéo (*cf.* image présente à droite du *workflow* présenté Figure 1).

Une présentation du prototype pourra être réalisée lors des journées Infox sur Seine aux personnes intéressées lors des échanges.

Perspectives

Le prototype présenté peut prétendre à un niveau de maturité technologique 4 (*TRL - Technology readiness level*), cependant pour améliorer cette maturité technologique et proposer une solution utilisable par les citoyens, impliqués professionnellement dans la vie politique ou non, il sera nécessaire d'appréhender et de lever plusieurs verrous scientifiques et technologiques :

- la nécessité d'enrichir la base de données d'énoncés politiques et des productions de *fact-checking* afférentes ;
- le renforcement de la robustesse et l'explicabilité des algorithmes de *matching* ;
- la considération de l'aspect temps-réel pour basculer d'un traitement pour l'heure *offline* à un traitement en ligne ;
- la gestion des situations multi-vocales (en particulier les débats, où la reconnaissance, la transcription et la détection sont rendues plus complexes).

² <https://dataset-stvd.univ-tours.fr/fc/data/factoscope.zip>

³ <https://dataset-stvd.univ-tours.fr/fc/>

⁴ <https://alphacephei.com/vosk/>

Par ailleurs, des études faisant suite à nos travaux pourront être menées :

- en matière d'interface homme-machine : penser l'apparition et le positionnement des *pop-up*, leur suivi, leur consultation, et le *feedback* des futurs utilisateurs ;
- en matière d'analyse du discours politique : compréhension des subtilités, des éléments de langage, des formulations rhétoriques, analyse des contradictions dans le temps, etc.

Références

Adair Bill, Chengkai Li, Jun Yang and Cong Yu. "Automated pop-up fact-checking: challenges and progress". Proceedings of the 2019 Computation+Journalism Symposium. 2019

Bigot Laurent, *Fact-checking vs. fake news. Vérifier pour mieux informer*, INA Editions. 2019

Graves Lucas, 2018, "Understanding the promise and limits of automated fact-checking", in *Reuters Institute for the Study of Journalism Factsheets*, Oxford University. 2019
<https://ora.ox.ac.uk/objects/uuid:f321ff43-05f0-4430-b978-f5f517b73b9b>

Kantar Public One Point / La Croix 2023 (23 janvier), « Le recours aux fact-checkers », in *La confiance des Français dans les médias. Etude Kantar Public One Point pour La Croix*.
<https://www.kantarpublic.com/fr/barometres/barometre-de-la-confiance-des-francais-dans-les-media>, p. 52

Nicey Jérémie, Bigot Laurent, « *Le soutien de Google et de Facebook au fact-checking français : entre transparence et dépendance* », Sur le journalisme, Vol. 9, n°1, pp. 188-203. 2020
<http://www.surlejournalisme.kinghost.net/rev/index.php/slj/article/view/417>

Pennycook Gordon, Bear Adam, Collins Evan T., Rand David G., "The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings", *Management Science*. 2019
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035384

Rayar Frédéric, Delalandre Mathieu, Le Van-Hao, "A large-scale TV video and metadata database for French political content analysis and fact-checking". International Conference on Content-based Multimedia Indexing. 2022

Roozenbeek Jan, van der Linden Sander, Goldberg Beth, Rathje Steve, Lewandowsky Stephan, 2022, "Psychological inoculation improves resilience against misinformation on social media", *Science Advances*, Vol. 8, n. 34. 2022. <https://www.science.org/doi/10.1126/sciadv.abo6254>

Sauvageau Florian, Thibault Simon, Trudel Pierre (dir.), *Les fausses nouvelles : nouveaux visages, nouveaux défis*, Presses de l'Université Laval. 2018

Schroff Florian, Kalenichenko Dmitry, Philbin James, "FaceNet: A Unified Embedding for Face Recognition and Clustering". *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2015

Smalley Seth, 2022 (18 août), « Meta won't comment on its plans to abandon CrowdTangle », *Poynter Institute*, <https://www.poynter.org/reporting-editing/2022/meta-wont-comment-on-its-plans-to-abandon-crowdtangle/>