



Stratégies d'influence et de dissimulation des discours de haine racialisant sur la plateforme X

Mario Laurent, Jérôme Ferret

► To cite this version:

Mario Laurent, Jérôme Ferret. Stratégies d'influence et de dissimulation des discours de haine racialisant sur la plateforme X. Troisième Journée d'études de l'Arcom : présentation des travaux des chercheurs sur les médias audiovisuels et numériques, Arcom (Autorité de régulation de la communication audiovisuelle et numérique), Nov 2024, Paris, France. <hal-04957257>

HAL Id: hal-04957257

<https://hal.science/hal-04957257v1>

Submitted on 19 Feb 2025

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.



Distributed under a Creative Commons CC BY-NC-SA 4.0 - Attribution - Non-commercial use - ShareAlike - International License

Proposition de contribution - 3^e Journée d'Etudes de l'ARCOM - 14 novembre 2024

« Stratégies discursives d'influence et de dissimulation des discours de haine racialisant sur la plateforme X. »

Mario Laurent

Docteur en Sciences du Langage, Université Toulouse Capitole, iDETCOM, Toulouse, France.

Post-Doctorant financé par l'ICO.

mario.laurent@ut-capitole.fr

Jérôme Ferret

Maitre de conférences HDR (EHESS) en Sociologie, Université Toulouse Capitole, iDETCOM, MSHS T (Unité CNRS), Toulouse, France.

<https://orcid.org/0000-0003-2080-7162>

jerome.ferret@ut-capitole.fr

Résumé de la présentation

Les travaux actuellement menés en sciences humaines et sociales dans le champ des dites *Digital Studies*, afin d'identifier, de comprendre et d'étudier les groupes radicaux en ligne se heurtent à de nombreuses difficultés. Premier obstacle : chaque plateforme d'échanges en ligne impose ses propres règles qui évoluent constamment en fonction des intérêts de leurs détenteurs et des lois qui les régissent. Deuxièmement, les usages et les stratégies discursives sont repensées en permanence par des communautés qui s'affrontent sur ces plateformes pour gagner de la réputation et de l'attention.

Prenant en compte ces deux contraintes structurelles et structurantes, nous nous intéressons plus particulièrement à la plateforme X, anciennement Twitter, sur laquelle nous avons commencé à étudier et à analyser les discours de haine dès 2018.

Nos objectifs sont multiples : nous cherchons d'abord à repérer et décrire les stratégies évolutives mises en place par les groupes radicaux sur X pour séduire leur audience, et plus particulièrement à identifier les **procédés discursifs employés**. Nous souhaitons d'autre part définir des critères permettant de catégoriser de la manière la plus précise possible ces différents procédés afin de faciliter les travaux de détection automatiques menés par nos collègues en informatique et en traitement automatique des langues (TAL). Notre projet vise donc à **comprendre** et **modéliser** les discours radicaux en ligne afin de mieux les **détecter** et les **réguler**.

Question de recherche

Quels sont les stratégies d'influence et les procédés discursifs employés sur X par les groupes radicaux ? Comment identifier et catégoriser ces procédés pour aider à la régulation des discours ?

Nous concentrerons notre présentation autour des **discours racialisant en ligne**. En effet, les groupes radicaux racistes font peser plusieurs dangers sur les utilisateurs des réseaux sociaux, à l'échelle individuelle comme à l'échelle collective. Leurs armes comportent la haine qui incite à la discrimination et à la violence contre certains groupes sociaux ; et la manipulation qui amène à la déstabilisation du système démocratique. La haine s'exprime principalement par la propagation de stéréotypes, ciblant des groupes entiers et par l'organisation de campagnes de harcèlement envers des ennemis désignés plus précisément. La manipulation fait notamment appel au partage de fausses informations ou d'informations formatées par un cadre idéologique. Ces différentes tactiques peuvent ensuite engendrer des actions violentes commises hors des réseaux sociaux.

Nous considérons que l'étude de la plateforme X reste pertinente en 2024 puisqu'il s'agit du réseau social le plus utilisé pour s'informer (Etude ARCOM, 2024). C'est toutefois une plateforme très mal modérée (voir le rapport de transparence de X pour le DSA en novembre 2023). **De plus, sur la plateforme X en particulier, la radicalité des discours peut être considérée comme la norme.** Elle en constitue même le moteur. La mécanique algorithmique joue ainsi un rôle en favorisant les publications clivantes et les *fake news* (Longhi, 2017 ; Vosoughi & al., 2018). Les utilisateurs sont ainsi naturellement polarisés et les discours constituent un terrain d'affrontement permanent sur les représentations. Ces représentations du monde sortent du champ scientifique et produisent une idéologisation du champ médiatique conventionnel (Cardon, 2019).

Données

Notre accès aux données de la plateforme X étant actuellement entravé par la politique actuelle de vente de leur API permettant d'accéder aux données exhaustives, nous appuierons nos analyses sur des données de projets précédents, notamment le corpus SoSweet, disponible en ligne sur la plateforme ORTOLANG, et nos propres données issues du projet Hatemeter. Nous illustrerons également l'évolution des stratégies discursives avec des exemples de publications sur X plus récents.

Stratégies d'influence et de dissimulation

Nous listerons d'abord les tactiques générales utilisées par des groupes racistes : (a) imposer un cadrage idéologique (notion de *framing* (Gamson & Modigliani, 1987)) ; (b) séduction de l'audience et normalisation (notamment la recours à la dimension para-sociale et à l'humour) ; (c) sophistication et dissimulation du langage (stéréotypes (Fiske, 1998), discours implicites (Schmeisser-Nieto et al. (2022)) et *dog whistles*).

Nous entrerons ensuite dans le détail des méthodes de dissimulation récurrentes sur X. Nous expliciterons plus précisément la tactique (c) et l'origine, l'évolution de plusieurs *dog whistles*. Nous verrons que certaines expressions, comme le « *Grand Remplacement* », peuvent prendre différents statuts au fur et à mesure de leur entrée et normalisation dans le langage commun.

Projet de typologisation des discours racialisant

Nos travaux encore en cours aboutiront à la structuration de plusieurs typologies concernant les différents niveaux des discours que nous étudions : Objet, Méthode, Outil, Cible. Ils sont amenés à être consolidés par l'étude quantitative d'autres données auxquelles nous devrions avoir accès par la suite. Nos typologies devraient permettre à la recherche d'améliorer les outils de détection automatique, en fournissant des manuels d'annotations d'énoncés implicites détaillés mais aussi à rédiger du matériel éducatif pour sensibiliser les utilisateurs. Nous avons des projets de formation à la fois auprès d'étudiants à l'université et de public adulte via la région Occitanie.

Références

- ARCOM (2024). *Les Français et l'information*. [<https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/les-francais-et-linformation>]
- Cardon, D. (2019). *Pourquoi avons-nous si peur des fake news ?*. Média en ligne, AOC. [<https://aoc.media/analyse/2019/06/20/pourquoi-avons-nous-si-peur-des-fake-news-1-2/>]
- Fiske, S., T. (1998). *Stereotyping, prejudice, and discrimination*, in *The handbook of social psychology*, D. T. Gilbert, S. T. Fiske, and G. Lindzey, Eds. McGraw-Hill, pp. 357– 411.
- Gamson, W., and Modigliani, A. (1987). *The Changing Culture of Affirmative Action*, in *Research in Political Sociology*, R. Braungart, Ed. Greenwich, CT, USA: JAI Press Inc, pp. 137– 177.
- ICAR - UMR 519, DANTE Inria, LIDILEM - EA 609, ALMANACH. (2024). *SoSweet [Corpus]*. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, v1. [<https://hdl.handle.net/11403/sosweet/v1>]
- Longhi, J. (2017). *Tweets politiques : corrélation entre forme linguistique et information véhiculée, #Info. Partag. Comment. l'info sur Twitter Faceb.*
- Rapport de transparence de X (2023). *Rapport de novembre 2023 en respect du Digital Service Act*. [<https://transparency.x.com/dsa-transparency-report-2023.html>]
- Schmeisser-Nieto, W., Nofre, M. and Taulé, M. (2022). *Criteria for the Annotation of Implicit Stereotypes*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 753–762, Marseille, France. European Language Resources Association.
- Vosoughi, S., Roy, D. and Aral., S. (2018). *The spread of true and false news online*, *Science* (80), vol. 359, no. 6380, pp. 1146–1151.