



**HAL**  
open science

## **La plateforme des médias français et le ton du débat public**

Inna Lyubareva, Julien Mésangeau, Nadira Boudjani, Imad El Badisy, Laurent Brisson

### ► **To cite this version:**

Inna Lyubareva, Julien Mésangeau, Nadira Boudjani, Imad El Badisy, Laurent Brisson. La plateforme des médias français et le ton du débat public. *Revue Communication (Collection Actes)*, 2021, 38 (2), <10.4000/communication.14433>. <hal-03373625>

**HAL Id: hal-03373625**

**<https://imt-atlantique.hal.science/hal-03373625v1>**

Submitted on 13 Dec 2022

**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-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

# La plateforme des médias français et le ton du débat public

Exemple de YouTube

Inna Lyubareva, Julien Mesangeau, Nadira Boudjani, Imad El Badisy et Laurent Brisson

---



## Édition électronique

URL : <https://journals.openedition.org/communication/14433>

DOI : [10.4000/communication.14433](https://doi.org/10.4000/communication.14433)

ISBN : 978-2-925203-01-8

ISSN : 1920-7344

## Éditeur

Université Laval

Ce document vous est offert par IMT Atlantique



## Référence électronique

Inna Lyubareva, Julien Mesangeau, Nadira Boudjani, Imad El Badisy et Laurent Brisson, « La plateforme des médias français et le ton du débat public », *Communication* [En ligne], Vol. 38/2 | 2021, mis en ligne le 01 novembre 2021, consulté le 13 décembre 2022. URL : <http://journals.openedition.org/communication/14433> ; DOI : <https://doi.org/10.4000/communication.14433>

---

Ce document a été généré automatiquement le 16 novembre 2021.



Creative Commons - Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International - CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

---

# La plateformes des médias français et le ton du débat public

Exemple de YouTube

Inna Lyubareva, Julien Mesangeau, Nadira Boudjani, Imad El Badisy et Laurent Brisson

---

- 1 Dans le contexte numérique, les médias sont souvent amenés à offrir leurs contenus aux consommateurs par l'intermédiaire des réseaux sociaux en ligne et autres plateformes numériques. Ce phénomène de plateformes s'accompagne de la formation de nouveaux espaces publics d'interaction, qui deviennent aujourd'hui de nouveaux lieux « d'exercice de la parole publique [...] de production et de circulation sociale du sens et [...] de débats relatifs à la mise en discours du social » (Delforce, 2010, p. 58). Dans ce contexte, les commentaires, *tweets* et autres *reviews* des internautes constituent une source de réintermédiation (Rebillard et Smyrniaos, 2010) de l'information journalistique : les médias traditionnels produisent une offre d'information conforme aux formats attendus sur les médias sociaux, tandis que les utilisateurs en viennent à occuper une position nouvelle, entre consommateurs et producteurs. Cette transformation du paysage informationnel est accompagnée de nombreux bouleversements.
- 2 Techniquement, la promesse d'une sphère publique participative, implicite, de Twitter, YouTube ou encore Facebook, est tenue : les individus disposent d'un cadre matériel susceptible d'héberger la « communauté de communication idéale » (Apel, 2001). Sociologiquement, en revanche, les participants n'affichent pas nécessairement l'objectif d'établir une intersubjectivité au travers de débats dont émergerait le consensus. En effet, de nombreux travaux mettent en évidence la « brutalisation du débat » (Badouard, 2017) et le déroulement d'affrontements polémiques au ton agressif (Tuomarla, 2014 ; Amossy et Burger, 2011) dans ces espaces. Il y a donc une « désinhibition toxique » des individus (Wachs et Wright, 2018) accompagnée souvent d'une incapacité des témoins à déterminer et à exercer une sanction (Neto *et al.*, 2017 ; Bastiaensens *et al.*, 2015). Plusieurs travaux démontrent aussi qu'une exposition croissante à la violence des utilisateurs des réseaux sociaux, d'une part, augmente le

manque d'estime de soi, la solitude, l'isolement social, la dépression, voire le suicide (Primack *et al.*, 2017 ; Luxton *et al.*, 2012 ; Vogel *et al.*, 2014 ; Bauman, 2013) et, d'autre part, pousse les utilisateurs eux-mêmes à agir de manière agressive (Hsueh, 2015 ; Rösner *et al.*, 2016) tout en favorisant l'émergence d'une méfiance envers les médias (Borah, 2013 ; Anderson *et al.*, 2016).

- 3 Les publications scientifiques lient souvent la brutalisation du débat aux facteurs psychologiques et sociaux (cf. Jane, 2015 pour une discussion critique des approches existantes). Elles interrogent également les relations complexes qui lient le design au fonctionnement des espaces en ligne dans une suite de productions soulignant les rôles joués par l'asynchronie de la communication, l'anonymisation des participations ou la tolérance à l'agressivité en ligne (Suler, 2004 ; Moor *et al.*, 2010 ; Halpern et Gibbs, 2013 ; Wright et Street, 2007). De plus, les médias, contrairement aux pratiques de modération plus strictes qu'ils mettent en œuvre sur leurs sites Web, peuvent tolérer un débat plus brutal et agressif sur les chaînes et comptes qu'ils créent sur les médias et réseaux sociaux ; en cause notamment, les objectifs de maximisation d'audience en ligne (Carlson, 2018 ; Humprecht *et al.*, 2020).
- 4 Malgré la littérature croissante dans ce domaine, rares sont les travaux qui offrent une lecture embrassant un large éventail de médias afin d'y analyser l'expression de l'agressivité ou qui proposent une méthodologie permettant de formaliser la relation entre l'agressivité de la participation, les caractéristiques des médias et leurs publics (Humprecht *et al.*, 2020). Dès lors que l'on prend les espaces francophones de commentaires des médias sociaux pour terrain, de tels travaux deviennent (quasi) inexistantes. Pour combler cette lacune, notre analyse se focalise sur la plateforme YouTube, souvent considérée comme particulièrement exposée aux phénomènes d'agressivité du débat par comparaison avec d'autres réseaux sociaux (Halpern et Gibbs, 2013).
- 5 Nous nous basons sur une étude inédite de 2 209 206 commentaires, dont 1 184 859 se trouvent dans les fils de discussion<sup>1</sup>, répartis dans les espaces de commentaires de 46 090 vidéos. Celles-ci sont issues d'un ensemble de 57 chaînes de médias français aux catégories institutionnelles et aux positionnements différents. Cet échantillon (cf. annexe 1) est en partie construit sur la base des travaux empiriques précédents (Marty *et al.*, 2012 ; Cardon *et al.*, 2019 ; Lyubareva et Rochelandet, 2016 ; Lyubareva *et al.*, 2020a), recensant des médias de presse traditionnels, nationaux et régionaux et les *pure players* (par exemple, Reporterre ou Slate). Parmi les médias cités dans ces recherches et représentatifs de leurs domaines, nous avons sélectionné 57 acteurs disposant d'une chaîne YouTube. Sans prétendre à une couverture exhaustive des médias français présents sur YouTube, cette sélection nous permet d'établir une catégorisation des chaînes en trois groupes (cf. annexe 1) : les médias généralistes couvrant un large éventail de thématiques, qui existent en version papier, en version numérique ou à la télévision, les médias « partisans » de gauche et de droite marqués par un positionnement politique plus extrême, qui ont une version papier ou qui existent uniquement en version numérique et les médias thématiques et spécialisés couvrant exclusivement certains sujets ou zones géographiques.
- 6 Nous nous posons pour objectif de comprendre, d'une part, où et quand l'espace des commentaires des médias français en vient à être touché par les débats brutaux ou agressifs et, d'autre part, quels facteurs limitent ou favorisent cette agressivité. Pour atteindre cet objectif, nous employons un dispositif d'enquête inhabituel au regard des

méthodes utilisées dans les travaux abordant le terrain YouTube en sciences sociales ; ce dispositif se positionne à la lisière du traitement automatisé du langage et de l'économie politique des médias (Guibert *et al.*, 2016).

- 7 Dans la première partie, nous allons résumer certaines orientations retrouvées dans la littérature afin de préciser quels facteurs sont susceptibles de peser sur l'agressivité dans les commentaires YouTube et de comprendre sous quelles conditions l'agressivité intervient dans les espaces de commentaires de la plateforme. Dans la deuxième partie, nous reviendrons sur les données et la méthodologie mise en œuvre. Nous verrons ainsi dans notre analyse que des formes de participation très différentes sont observées selon que nous considérons les commentaires sur les chaînes de médias que nous classons comme « généralistes », ceux des médias que nous classons comme « partisans » ou enfin ceux des médias que nous classons comme « thématiques et spécialisés ».
- 8 Dans la troisième partie, nous parcourons les résultats de cette recherche et vérifierons comment les médias en ligne, espaces de commentaires et d'agressivité sont liés par un ensemble de pratiques oscillant, selon les chaînes et les niveaux de fréquentation des espaces de commentaires des vidéos, de peu à fortement agressives.

## Revue de la littérature

- 9 Dès 2008, les sciences humaines et sociales interrogeaient la capacité de YouTube à offrir à l'« opinion publique connectée » (Vanbremeersch, 2009) un espace préservant le pluralisme. Edgerly *et al.* observent au sujet de YouTube que la section de commentaires est un espace « qui a le potentiel de réaliser le type de discussion publique qui a longtemps été valorisé comme une partie importante de la démocratie » (2009, p. 3). Pour ces auteurs, YouTube permet la création de chaînes dotées d'une orientation marquée (sujets de société ; commentaire d'œuvres culturelles ; journalisme et information) et offre donc à des commentateurs la possibilité de trouver des sujets d'intérêt, de se réapproprier les contenus et, à terme, de participer au débat. Ils notent également que si « YouTube offre une arène délibérative pour les conversations politiques en ligne, [la plateforme] ne rencontre pas les standards minimaux de la sphère publique » (*ibid.*, p. 6). Kwon et Cho (2017) apportent, près d'une décennie plus tard, une lecture sensiblement similaire. Pour ces derniers, YouTube est un exemple d'espace public particulièrement touché par des conflits et des discours agressifs. En observant la façon dont les commentateurs participent aux fils de commentaires des vidéos de campagne de Donald Trump, les auteurs ont expliqué ce phénomène d'agressivité par un effet domino engagé sitôt le propos injurieux mis en ligne. Halpern et Gibbs (2013), dans leur analyse comparative des plateformes YouTube et Facebook, constatent une agressivité plus élevée de la première et lient ce résultat à une propension plus marquée des participants à conserver un anonymat ainsi qu'à une désindividualisation largement favorisée par la plateforme. Nous pouvons ainsi supposer que les normes d'un espace de commentaires tiennent tant qu'une transgression n'a pas lieu. Sitôt celle-ci réalisée, l'espace va très vite dériver vers plus d'insultes et de haine. Reprenant le constat des auteurs, nous nous proposons d'analyser quels sont les contextes sur YouTube plus et moins favorables à l'émergence de tels comportements.

## Homogénéité des publics et agressivité : les médias partisans et spécialisés à l'abri des propos agressifs ?

- 10 Comme le note Charon, l'information et sa distribution ont évolué, avec Internet, vers une personnalisation permettant aux offres de se décliner en « propositions toujours plus spécialisées » (Charon, 2015, p.88). Ce phénomène s'inscrit dans une tendance de longue durée, déjà présente avec la radio FM et la télévision qui proposaient des « découp[es] des communautés de publics toujours plus fines » (*ibid.*). Sur les médias sociaux, et sur YouTube en particulier, cette évolution offre aux médias divers moyens d'action, dans les pratiques d'éditorialisation (dans les vidéos, leurs thèmes, la diversité des positions représentées) et le cadrage de la réception (les débats et remarques de commentateurs).
- 11 Certains médias font converger leurs productions sur les différents médias, ce qui amène une chaîne YouTube à couvrir la même diversité de sujets que la version numérique ou papier d'un grand quotidien de la presse nationale par exemple. *Le Monde* est, à ce titre, un bon exemple. Dans ce type de configuration, la grande diversité des sujets traités et des angles adoptés va créer, pour reprendre les mots de Remy, un YouTube d'« extrême hétérogénéité » où « la section de commentaires d'une vidéo YouTube se présente comme un patchwork de débats très divers [...] souvent déconnectés les uns des autres et parfois même de la vidéo sous laquelle ils ont été intégrés » (2018, p. 71). Cette grande hétérogénéité est finalement ce qui confère aussi à une chaîne YouTube sa capacité à endosser les attributs de la sphère publique, une sphère qui est, pour reprendre Habermas, « enracinée dans des réseaux de flux de messages regroupant news, commentaires, discussions, scènes et images, émissions et films avec une visée informative, polémique, éducationnelle ou de divertissement » (2006, p. 415). Cette diversité autorise et favorise la constitution d'une audience plurielle, dans ses goûts et ses valeurs.
- 12 À l'opposé des médias généralistes offrant un espace de discussion pluraliste, les médias que nous classerons en « partisans » reprennent plus volontiers une ligne éditoriale resserrée, avec un traitement de l'information par l'opinion plus présent – à l'image des vidéos d'Usul sur *Mediapart*. Finalement, nous pouvons distinguer des médias « spécialisés et thématiques » où les utilisateurs sont regroupés en raison du partage d'un intérêt commun. À l'instar des « tiers-lieu » (Oldenburg, 1999), les espaces de discussion en ligne associés à ces médias séparent un espace public offrant aux personnes d'opinions différentes un lieu de discussion et de délibération « domestique » hébergeant les conversations entre personnes partageant les mêmes sensibilités et attirant un public régulier, dans un cadre excluant les positions et les discussions clivantes (Wright, 2016 ; Wright *et al.*, 2016).
- 13 En fonction du positionnement des médias et du public qu'ils visent, la littérature formule certaines prédictions au sujet du niveau d'agressivité dans les espaces de commentaires anglophones des médias sociaux et réseaux socionumériques.
- 14 Une analyse comparative récente (Humprecht *et al.*, 2020), portant sur les espaces de commentaires sur Facebook de plusieurs médias américains et allemands, met en lumière que les médias généralistes *mainstream* accueillent davantage d'insultes et de haine que les médias plus spécialisés haut de gamme. Les auteurs expliquent ce phénomène, d'une part, par la présence dans les espaces de commentaires de ces médias d'un public plus hétérogène et, d'autre part, par le fonctionnement même de

leurs modèles économiques. À propos de ce dernier élément, les recherches (Humprecht et Esser, 2018) montrent que les médias généralistes orientés grand public utilisent les réseaux sociaux numériques dans l'objectif d'attirer l'audience supplémentaire la plus large possible (en plus de leurs autres canaux de distribution) et de maximiser leurs revenus. Or, il a été démontré que la haine et la provocation dans les commentaires attirent plus d'audience, stimulent son attention, et permettent donc à une information d'atteindre un public plus large (Muddiman et Stroud, 2017 ; Weber, 2014 ; Ziegele *et al.*, 2018). Par conséquent, les médias généralistes pourraient être plus enclins à tolérer les commentaires agressifs dans leurs espaces de discussion sur les réseaux sociaux numériques.

- 15 À côté, la catégorie regroupant les médias « spécialisés et thématiques », médias couvrant exclusivement certains sujets ou zones géographiques, repose sur une homogénéité d'un public. Leurs espaces de commentaires sont habités de conversations regroupant des publics qui partagent des intérêts et engagements communs couvrant divers domaines de la vie sociale. À ce titre, des individus de diverses orientations et sensibilités sont amenés à échanger dans un cadre où sont discutés des événements transcendant le cadre des adhésions politiques partisans (Barbera *et al.*, 2015). Dans ces espaces, les utilisateurs vont, au fil de leurs commentaires, implicitement participer à la définition de principes de communication favorisant la solidarité et limitant les comportements déviants, la toxicité d'un *trolling* notamment (Graham, 2012 ; Wright *et al.*, 2016).
- 16 Finalement, les travaux sur les médias partisans prédisent des résultats divergents en matière d'agressivité sur les plateformes en ligne.
- 17 Le premier groupe de travaux met en avant le phénomène de polarisation particulièrement présent sur les espaces partisans de discussion en ligne (Davis, 1999 ; Wright *et al.*, 2016). Il se traduit par la réduction progressive des orientations et thèmes susceptibles d'être rencontrés dans les vidéos consommées et les espaces commentés par les utilisateurs de réseaux sociaux numériques. Le fonctionnement des algorithmes de recommandation des plateformes renforce ce phénomène.
- 18 Historiquement, les systèmes de recommandation tendent à privilégier un principe de filtrage basé sur le contenu, où l'utilisateur se voit proposer des éléments semblables à ceux consommés par le passé (Adomavicius et Tuzhilin, 2005). Ils reposent donc sur l'identification des centres d'intérêt des utilisateurs et proposent à ces derniers des contenus similaires (Lops *et al.*, 2011). Comme en témoigne la revue de littérature de Peis *et al.* (2008), durant la décennie 2000-2010, les algorithmes tendent à recommander des contenus sans que les avis et pratiques des autres utilisateurs des plateformes concernées soient nécessairement exploités pour orienter la consommation d'un utilisateur. À l'instar de certains concurrents, YouTube va innover sur ce point. À l'occasion d'un colloque scientifique, plusieurs ingénieurs employés par la plateforme ont offert un aperçu du fonctionnement de son algorithme de recommandation (Covington *et al.*, 2016). Nous apprenons que l'algorithme de recommandation de YouTube se base sur les comportements passés des utilisateurs pour définir des recommandations communes à ceux qui partagent un même historique d'activité — ce que les ingénieurs à l'origine de l'algorithme ont nommé le « filtrage collaboratif ». En somme, les recommandations vont désormais exploiter des données d'usages mutualisés, ceux de l'utilisateur A servant à recommander l'utilisateur B, et ainsi de suite. Incidemment, plus les individus tendent à favoriser la consultation partisane de

vidéos ancrées à gauche ou à droite, avec une ligne éditoriale forte, plus les contenus proposés par YouTube seront affinés et conformes aux aspirations (et consultations) premières des individus, plus le public fidèle à ces contenus sera homogène. À l'exception des utilisateurs qui feront le choix délibéré d'ignorer les recommandations offertes par la plateforme (Bakshy *et al.*, 2015), cette réduction progressive de la diversité des contenus va diminuer les risques, pour les utilisateurs, d'être confrontés à des opinions et positions non conformes à leurs valeurs (Sunstein, 2017). Par extension, dans la lignée des observations posées par Cass Sunstein (2011, 2017), nous pouvons présumer que cette uniformisation des thèmes et orientations — principe structurant et central des médias sociaux — est susceptible de diminuer l'agressivité des débats : puisque les individus, de consultation de vidéo en consultation de vidéo, réduisent les risques d'être confrontés à des personnes ne partageant pas leurs appétences, le risque d'être engagés dans un débat où les échanges deviennent agressifs diminue.

- 19 Ainsi, les médias partisans constituent des espaces de prise de parole fréquentés par des publics partageant — *a minima* — une certaine sensibilité politique, voire une forme plus ou moins marquée de compagnonnage basée sur le partage de positions politiques radicales. Au-delà des préconisations de contenus aux individus, l'algorithme va aussi aboutir au renforcement d'une homophilie de valeur (Lazarsfeld et Merton, 1954). En effet, le public retrouvé dans l'espace de commentaires — parce qu'il partage des valeurs, des historiques de consultation de contenus en partie semblables et évolue dans l'espace de médias partisans — sera un terreau favorable à l'émergence d'un contrôle social, qui justifie la sanction des individus selon qu'ils communiquent — ou pas — d'une manière jugée acceptable par leurs pairs. Comme le précisent Marques *et al.* (2001), un groupe formé par une adhésion commune à un ensemble de valeurs, ou encore à une identité, tend à défendre l'uniformité en son sein, cette pression normative ayant pour rôle de maintenir une définition durable des croyances et opinions du groupe et limitant la présence de comportements agressifs et toxiques.
- 20 Contrairement à ces premiers résultats, d'autres travaux mettent en lumière le fait que ces zones de débat fortement polarisées, si elles sont effectivement idéologiquement plus homogènes, restent traversées par des tensions internes caractérisées par la prise de parole d'ardents opposants.
- 21 Ces derniers utilisent les zones de commentaires pour exprimer, souvent sous forme brutale et toxique, leur désaccord avec l'opinion collective (Humprecht *et al.*, 2020). En retour, ces commentaires négatifs provoquent des réactions offensives de la part de majorité. Ce phénomène a été aussi observé dans une étude récente sur la sphère des discours masculinistes sur YouTube où un « contre-public » fut identifié. Ce contre-public est composé de commentateurs faisant front autour des discours autorisés sur l'homme et le genre, et qui s'opposent agressivement — et collectivement — à ceux portent un discours contradictoire, notamment en ayant recours à l'agression et à la dérision (Mesangeau et Morin, 2021).
- 22 Ainsi, dans les espaces de commentaires très marqués politiquement, caractéristiques des médias partisans, avoir l'ascendance lors d'un débat devient structurant, contrairement à ce que prescrivent les normes de solidarité et réciprocité régulant la participation dans des espaces non explicitement politiques ou idéologiques (Wright *et al.*, 2016 ; Graham, 2012). Ces éléments permettent de suggérer que la polarisation et l'agressivité peuvent être fortement corrélées dans les espaces partisans de commentaires (Miller et Conover, 2015).

- 23 Pour résumer, et en nous basant sur la revue de la littérature présentée ci-dessus, nous présumons, et c'est notre hypothèse 1, que le débat agressif sera moins présent dans les chaînes que nous classons dans la catégorie des médias « spécialisés et thématiques », notamment car ces chaînes occupent, contrairement aux médias « généralistes » et « partisans », des sphères éditoriales très précises, liées à une spécialisation thématique et qui n'est pas explicitement politique ou idéologique.

### L'agressivité, produit de la quantité ?

- 24 Schultes *et al.* (2013) observent que l'agressivité pèse fortement sur les perceptions et pratiques des utilisateurs de YouTube. Cette observation est également vérifiée dans les espaces de commentaires des vidéos d'actualité publiées par les médias (Khan et Jacob, 2013). Au terme d'une analyse portant sur les pratiques de commentaire sur YouTube — et en particulier sur le lien unissant anonymat et tendance à l'insulte (*flaming*) —, Khan et Jacob rejoignent les conclusions de Moor *et al.* (2010) ainsi que de Halpern et Gibbs (2013). Ils observent que le *flaming* est commun sur YouTube et qu'il est systématiquement issu de commentateurs anonymes. Suivant ces auteurs, nous pouvons retenir que plus l'espace de commentaires permet l'anonymat, plus nous y retrouvons de commentateurs empruntant un registre de conversation violent et agressif. Or, sur les médias sociaux, l'anonymat n'est pas seulement affaire de pseudonyme. De façon générale, la rencontre violente de publics diversifiés est favorisée par le contexte de la communication médiatisée (Barnes, 2003 ; Reicher *et al.*, 1995). Au-delà, comme le notait Badouard (2017), la communication sur un réseau socionumérique, ou un média social dans le cas de YouTube, va fréquemment dévier vers des échanges agressifs, les principaux responsables étant ici — comme nous l'évoquions en introduction — les possibilités d'anonymat et la désindividuation des échanges. Au-delà de la violence banalisée de la conversation qui dérape, le média social favorise aussi un changement d'échelle de ces tensions et conflits. Kwon et Gruzd (2017) évoquent à ce titre la nature contagieuse de l'agressivité et des comportements mimétiques. Badouard et plusieurs autres chercheurs partagent aussi l'idée que cette agressivité est privilégiée par certains comme moyen de se faire entendre (Badouard, 2017 ; Vincent et Barbeau, 2012 ; Culpeper, 2011). Sur YouTube, à la suite de ces auteurs, nous considérons que deux facteurs vont créer les conditions favorisant un anonymat réel des individus, et ainsi limiter les possibilités de modération des propos agressifs / déviants par les participants.
- 25 Le premier facteur est retrouvé lorsque les individus de passage ne partagent ni position, ni sensibilité, ni identité avec les autres commentateurs. Ce manque d'attributs communs va en retour limiter l'apparition et l'expression du contrôle social que nous avons évoquées dans notre première hypothèse, contrôle qui permettrait pourtant au collectif de contenir les conduites inappropriées (par le signalement, la modération directe de la chaîne ou l'exercice d'une sanction sociale). Nous sommes alors confrontés à la situation opposée à celle décrite par Marques *et al.* (2001). Comme discuté dans l'argument menant à notre première hypothèse, de telles logiques sont normalement basées sur une intersubjectivité — à défaut d'une interconnaissance — issue du partage des mêmes pratiques de consultation de vidéos et de la fréquentation des mêmes espaces de commentaires. Sans ces logiques sociales, les commentateurs

perdent une partie des moyens à leur disposition pour réguler les conduites de leurs pairs.

- 26 Le second facteur est la quantité de commentaires et de discussions, signes de l'attention importante portée par les utilisateurs à une vidéo ou à une chaîne. Construisant notre proposition sur les observations de Badouard (2017) et Kwon et Gruz (2017), nous supposons que plus l'activité est importante, plus il sera difficile pour l'auteur de la chaîne de contrôler la tonalité des échanges, et pour les commentateurs de coordonner une sanction envers une personne aux propos agressifs. En conséquence, l'agressivité, renforcée par la grande visibilité des échanges ayant lieu sur des espaces fortement commentés (Pfeffer *et al.*, 2014), peut atteindre une échelle particulièrement importante compte tenu, d'une part, des possibilités limitées de sanction et, d'autre part, de la perception positive que les individus peuvent avoir de l'agressivité dans certaines conditions (Moor *et al.*, 2010).
- 27 Partant, nous proposons pour hypothèse 2 que les espaces de commentaires très fréquentés, marqués par une forte activité, en nombre de commentaires, de discussions et de participants, hébergeront plus de propos agressifs.

### Dimension communautaire de la plateforme et agressivité

- 28 Une littérature abondante a été développée sur le fonctionnement des groupes en ligne. Ces travaux soulignent le rôle des plateformes qui favorisent l'émergence de communautés d'utilisateurs basées sur les interactions répétitives, un principe d'homophilie et des intérêts partagés (Rheingold, 2000 ; Von Hippel, 2005 ; McPherson *et al.*, 2001 ; Cohendet *et al.*, 2003). Leurs principes structuraux sont aussi bien connus :
- L'émergence des groupes est le résultat d'interactions répétitives, qui se mettent en place entre les individus pour discuter, échanger de l'information et des ressources, apprendre ou jouer (Kraut et Resnick, 2012 ; Von Hippel, 2005) ;
  - Le principe d'homophilie, qui se trouve à l'origine des interactions sociales, se base dans le contexte numérique sur les intérêts partagés (en l'absence des facteurs sociodémographiques traditionnels) (McPherson *et al.*, 2001) ;
  - Il existe une corrélation positive entre l'activité de contribution d'un membre à l'activité du groupe et l'intensité de ses liens sociaux avec les autres membres (Laine *et al.*, 2011).
- 29 À l'image des autres plateformes (Lyubareva *et al.*, 2020b), YouTube pourrait également être à l'origine des communautés des utilisateurs qui « se rencontrent » régulièrement dans les espaces de commentaires des différentes chaînes. Certains travaux récents (cf. par exemple Georges, 2019) mettent en lumière, principalement avec des méthodes qualitatives, l'existence de telles communautés transversales sur YouTube.
- 30 Le présent article ne vise pas l'analyse détaillée des communautés de la plateforme YouTube. Nous nous appuyons sur la littérature précédente pour examiner en quoi la dimension communautaire de certaines chaînes de médias peut, du fait de publics engagés et actifs partageant les mêmes intérêts, diminuer le risque qu'une discussion dégénère vers de l'agression verbale.
- 31 La dernière hypothèse que nous proposons de tester afin de mieux comprendre les mécanismes de l'agressivité sur les espaces de commentaires YouTube est dérivée des précédentes. Avec nos deux premières hypothèses, nous testons, d'une part, l'idée que les médias spécialisés vont drainer une activité issue d'acteurs homogènes qui limitent

le risque de débats agressifs et, d'autre part, celle que les médias très fréquentés sont structurellement propices à l'agressivité. Avec ces hypothèses, nous n'interrogeons pas directement les effets que la présence de certains individus, identifiés en raison d'attributs spécifiques, peut avoir sur l'expression de propos agressifs. Afin d'explorer ce dernier point, nous proposons de tester une troisième hypothèse. Ladhari *et al.* (2020) notaient que les youtubeurs populaires partagent avec leur public un ensemble d'attitudes et de valeurs. Comme présenté ci-dessus, l'homophilie du public, partageant ces valeurs et attitudes, est une condition corrélée aux échanges non agressifs entre les utilisateurs. Notre hypothèse est donc que la participation des commentateurs actifs et populaires sur YouTube, qui commentent plusieurs chaînes et dont les commentaires sont très appréciés par les autres en nombre de *likes*, va s'articuler avec une moindre agressivité dans les discussions.

## Méthodologie

- 32 L'analyse des hypothèses de cette étude requiert une méthodologie appropriée. Plusieurs travaux ont été proposés pour repérer l'agressivité dans les messages sur les réseaux sociaux. Une partie de ces travaux repose sur un classement des lexiques utilisés dans les commentaires et opposant des termes considérés comme « positifs » ou « négatifs ». Une telle technique rencontre deux limites importantes dans le contexte du langage, le plus souvent informel et familier, employé sur les médias sociaux. Premièrement, l'orthographe et la grammaire ne sont souvent pas respectées par les utilisateurs, ce qui rend difficile l'exploration automatique des tonalités d'opinion exprimées dans les commentaires. Deuxièmement, « [ces techniques] rencontrent leurs limites lorsqu'il s'agit d'appréhender des constructions syntaxiques complexes ou l'usage de certaines figures de style telles l'ironie ou le sarcasme, très présentes sur [les réseaux sociaux] » (Boyadjian et Velcin, 2017, p. 115).
- 33 Pour pallier ces problèmes, d'autres travaux, réalisés sur des corpus en anglais, proposent des approches plus élaborées combinant les caractéristiques linguistiques, lexicales et syntaxiques des commentaires (Nobata *et al.*, 2016 ; Brody et Diakopoulos, 2011 ; Chen *et al.*, 2012). L'idée de ces travaux est de proposer une façon de représenter les messages par un ensemble d'attributs qui sépare, de façon précise, les messages agressifs des messages non agressifs.
- 34 En lien avec ces travaux, nous proposons de reprendre les critères de pertinence normalement utilisés dans la détection de commentaires agressifs en anglais, notamment les caractéristiques lexicales et linguistiques. Cependant, pour adapter la démarche proposée dans la littérature à un corpus en langue française, une analyse qualitative préalable est nécessaire. Nous combinons ensuite cette analyse qualitative de commentaires avec l'analyse automatisée du corpus et un traitement statistique de celui-ci. Ces différentes étapes de notre travail sont présentées ci-dessous.
- 35 Lors de la première étape, nous avons collecté les commentaires des vidéos de 57 chaînes de notre échantillon. Pour l'ensemble des chaînes, nous avons collecté tous les commentaires publiés de la date de création de chaque chaîne à juillet 2019 en utilisant YouTube *Data API*. Afin de pouvoir catégoriser les commentaires, nous avons d'abord analysé manuellement plus de 2 000 commentaires de façon à identifier la façon dont l'agressivité y est exprimée. Avec cette approche, nous adoptons une

position méthodologique forte : l'analyse automatisée des discussions doit être subordonnée à une enquête qualitative préalable, indispensable pour intégrer les particularités du YouTube francophone.

36 Nous avons ainsi recherché des attributs qu'un programme pourrait ensuite discriminer et extraire automatiquement (Lewis, 1992). La démarche qualitative préalable, combinée avec les listes existantes de gros mots<sup>2</sup>, nous a permis de construire un répertoire d'attributs d'agressivité regroupant les insultes (mots, acronymes, expressions et émoticônes vulgaires) et les indices pragmatiques renseignant sur le ton agressif du texte (points d'exclamation et d'interrogation répétés, mots allongés, mots en majuscules). Plus précisément, chaque commentaire a été représenté par les attributs suivants<sup>3</sup> :

- Nombre d'occurrences de mots et d'expressions appartenant au lexique d'insultes ;
- Ponctuation qui ajoute, selon nos observations, un effet offensif aux commentaires. Un texte agressif peut contenir un ou plusieurs points d'exclamation ou d'interrogation. Par exemple, « tu es con !!! » apparaît davantage comme commentaire agressif que « tu es con ». La ponctuation accentue donc le ton agressif et irrespectueux comme l'illustrent ces commentaires extraits de notre corpus : « C'est bien que les gilets jaunes soit éborgnés ou tués ? POURRITURE !!!! » ; « les gilets jaunes ??? plutôt les casseurs qui veulent pas travailler » ;
- Nombre de mots en majuscules : les utilisateurs peuvent écrire des commentaires partiellement ou complètement en majuscules afin de renforcer le sens des mots. Par exemple : « tu es CON » est plus fort que « tu es con » ou « C'est toi qui n'a aucun humour GROS FILS DE PUTE » ou encore « ferme vos sale GUELE dac ? ». Dans la même lignée, les commentaires écrits en majuscules peuvent servir à exprimer le ton irrespectueux, ironique ou péjoratif : « BLABLABLA FEMME EN DANGER BLABLABLA FEMME QUI SOUFFRE BLABLABLA EGALITER FEMME HOMME BLABLABLA COMMENT TOUT RAMENER A LA CONDITION FEMININE LA FEMME VA DEVENIR LA PIRE PLAIE QUE LA TERRE EST CONNU » ;
- Nombre de mots allongés (caractères répétés séquentiellement au moins trois fois) : les mots allongés sont utilisés pour rendre un terme plus insultant ou plus offensif. Exemple : « tu es coooooon » est plus fort que « tu es con » ; ou encore un autre commentaire : « MDRRRRRR PENDEZ LES,CES JAMBONS ».

37 Pour pouvoir extraire automatiquement ces attributs à partir des commentaires, nous avons appliqué les prétraitements traditionnels (Hemalatha *et al.*, 2012) : suppression des adresses courriels ; suppression des hyperliens ; remplacement des mots allongés par leur écriture correcte ; suppression des signes diacritiques (accents, tréma, cédille) ; lemmatisation (remplacer un verbe par son infinitif, un adjectif par son masculin-singulier, etc.) ; suppression des chiffres.

38 À l'issue de cette étape, nous avons construit une base de données par vidéo combinant les informations sur l'origine de la vidéo (la chaîne et son appartenance à l'une des catégories « Média généraliste », « Média spécialisé » ou « Média partisan ») et les attributs d'agressivité (en nombre et en valeur moyenne par commentaire de chaque vidéo). En lien avec les travaux existants, nous suggérons qu'un débat (discussion) associé à une vidéo est constitué de la somme de ses commentaires. Ainsi, l'agressivité des commentaires individuels détermine le niveau global du débat associé à une vidéo (cf. ci-dessous le « score d'agressivité »).

- 39 Les caractéristiques de débat par vidéo, en lien avec nos hypothèses formulées, ont été ajoutées à cette base de données :
- Le nombre de commentaires ;
  - La moyenne des mots dans les commentaires (longueur) ;
  - Le nombre de fils de discussion après la vidéo : les commentaires « *top-level* » suivis d'au moins une réponse<sup>4</sup> ;
  - Le nombre de commentateurs uniques ;
  - Le nombre de commentateurs populaires dont les commentaires sont les plus *likés* sur la plateforme : les membres avec le nombre de *likes* supérieur à la moyenne sur la plateforme ;
  - Le nombre de commentateurs actifs, participant aux discussions sur plus de trois chaînes différentes.
- 40 À l'étape suivante, sur la base de ces données et à l'aide de la méthode de clustering *k-means* (Clausen, 2008 ; Hair *et al.*, 1995), nous avons mis en évidence l'existence de trois clusters de vidéos. Cette méthode consiste à définir des groupes homogènes de sorte que la variation totale intra-groupe soit réduite au minimum. Le but de cette étape est de tester la consistance des attributs d'agressivité<sup>5</sup>.
- 41 Finalement, nous construisons la variable d'intérêt à expliquer le « score d'agressivité » d'une vidéo, qui est une variable continue. Elle caractérise le niveau général d'agressivité de débat associé à une vidéo et prend des valeurs comprises entre 0 et 1. Cette variable a été construite à partir d'une pondération des valeurs des différents attributs d'agressivité que nous avons pu identifier manuellement dans les commentaires.
- 42 Plus précisément, cette variable est construite à partir d'une combinaison pondérée des différents attributs d'agressivité :
- 43  $\text{Score d'agressivité d'une vidéo} = 0,5 * (\text{nbre de mots d'insulte par commentaire}) + 0,125 * (\text{nbre de mots allongés par commentaire}) + 0,125 * (\text{nbre de mots en majuscules par commentaire}) + 0,125 * (\text{nbre d'exclamations par commentaire}) + 0,125 * (\text{nbre d'interrogations par commentaire})$
- 44 Il est important de préciser que les poids des valeurs des attributs d'agressivité ont été déterminés de manière subjective sur la base de nos observations de l'analyse qualitative. Ainsi, pour évaluer le niveau d'agressivité d'une discussion associée à une vidéo, nous attribuons un poids plus important (0,5) au nombre de mots d'insultes dans les commentaires et les poids moins importants (0,125) aux autres attributs d'agressivité qui nous semblent secondaires.
- 45 Finalement, pour modéliser la relation entre le score d'agressivité (variable à expliquer) et les caractéristiques de débat par vidéo (variables explicatives, formulées en lien avec nos hypothèses : catégorie de la chaîne, nombre de commentaires, etc.), nous avons utilisé un modèle de beta régression. Il s'agit d'une classe de modèle propre aux variables d'intérêt qui prennent des valeurs dans l'intervalle [0,1]. Cette approche intègre naturellement des caractéristiques telles que l'hétéroscédasticité (variances différentes des termes des résidus pour les variables explicatives) ou l'asymétrie de la distribution de la variable d'intérêt, qui sont couramment observées dans les données prenant des valeurs dans un tel intervalle, telles que les taux ou les proportions (Espinheira *et al.*, 2008 ; Zeileis *et al.*, 2010). Ainsi, les modèles de régression nous permettent d'analyser la relation de causalité entre les éléments du contexte présentés dans les hypothèses (catégorie de la chaîne, quantité d'interactions, présence des membres communautaires) et le score d'agressivité dans les vidéos<sup>6</sup>.

## Résultats

### Classification des vidéos YouTube

46 Nous remarquons tout d'abord qu'une sous-population assez importante de vidéos — 9 275 vidéos (20 %) — dans notre échantillon ne contient aucun attribut d'agressivité dans l'espace des commentaires. Cette population inclut 19,5 % des médias généralistes, 13,1 % des médias partisans (mis à part *Atlantico* et *Fdesouche* : 100 % de commentaires dans leurs vidéos contiennent des attributs d'agressivité, ce qui peut partiellement s'expliquer par une faible présence dans notre échantillon) et 28,4 % des médias spécialisés. Le tableau 1 présente, pour chaque catégorie des médias, une sélection de cinq chaînes YouTube avec le pourcentage le plus élevé des vidéos sans agressivité dans les commentaires. Les médias les plus touchés par l'agressivité dans les commentaires sont *Le Monde*, *LeHuffPost* et *Metronews* dans la catégorie « généralistes » (avec seulement 5,2 %, 11,5 % et 13,3 % respectivement des vidéos sans aucun attribut d'agressivité) ; *Fakirpresse*, *Mediapart* et *AgoraVox* dans la catégorie « partisans » (2,6 %, 8,3 % et 9,5 % respectivement) ; et *Orient XXI*, *StreetPress* et *Famille Chrétienne* dans la catégorie « spécialisés » (10,3 %, 17,5 %, et 22,7 % respectivement). De manière générale, les médias spécialisés semblent être globalement moins concernés par le problème de l'agressivité : maximum 50 % des vidéos sans agressivité et minimum 10,03 % dans le cas d'*Orient XXI*.

Tableau 1. Les chaînes médias et le pourcentage de leurs vidéos sans aucun attribut d'agressivité

Chaînes média	Vidéo sans agressivité	Vidéo total dans l'échantillon	%
<b>Généralistes : 7 110 vidéos sans agressivité</b>			
<i>DNA - Dernières Nouvelles d'Alsace</i>	6	10	60,0
<i>La Voix du Nord</i>	18	34	52,9
<i>Ouest-France</i>	443	994	44,6
<i>France Net Infos</i>	12	27	44,4
<i>Le Dauphiné Libéré</i>	87	219	39,7
<b>Partisanes : 486 vidéos sans agressivité</b>			
<i>Valeurs Actuelles</i>	37	111	33,4
<i>Lundi Matin</i>	22	89	24,7
<i>Lutte ouvrière</i>	7	30	23,3
<i>L'Humanité</i>	160	799	20,0
<i>Alternatives Économiques</i>	6	31	19,4
<b>Spécialisées : 1 679 vidéos sans agressivité</b>			
<i>ZDNet.fr</i>	1	2	50,0
<i>Corse-Matin Presse</i>	48	107	44,9
<i>Corse Net Infos</i>	42	111	37,8
<i>Gala</i>	53	170	31,2
<i>La Manche Libre</i>	17	55	30,9

47 Les médias dont les commentaires sont les plus exposés à l'agressivité sont *Fdesouche* et *Atlantico* (l'agressivité est présente dans les espaces de commentaires de 100 % de ses

vidéos), *Le Monde* (94,8 %), *Mediapart* (91,7 %) ou *AgoraVox* (90,5 %). Ces acteurs font partie des catégories « partisan » et « généralistes ».

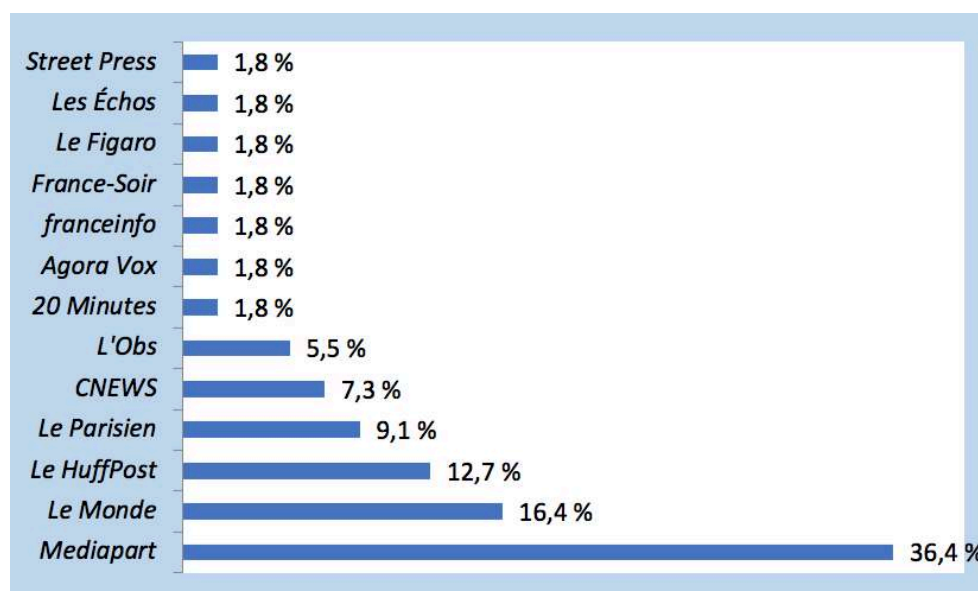
- 48 Une étape préliminaire de notre analyse vise à établir une classification de vidéos analysées en matière d'attributs d'agressivité retrouvés dans les commentaires. Les tableaux 2 et 3 présentent ces résultats.

Tableau 2. Classification des vidéos, valeurs moyennes par clusters (*k-means*)

Cluster	N <sup>bre</sup> de mots d'insulte	N <sup>bre</sup> de mots allongés	N <sup>bre</sup> de mots en majuscules	N <sup>bre</sup> d'exclamations	N <sup>bre</sup> d'interrogations
Cluster 1 : 55 vidéos	941,20	319,11	1 884,56	1 625,21	1 249,96
Cluster 2 : 755 vidéos	201,02	91,88	468,16	463,51	302,15
Cluster 3 : 45 280 vidéos	5,09	2,13	12,22	12,51	7,33

- 49 Le clustering fait apparaître trois clusters de vidéos, très déséquilibrés en matière de nombre de vidéos. Les clusters se distinguent significativement en matière d'attributs d'agressivité : par exemple, en moyenne, par vidéo, le cluster 1 contient 941,2 mots d'insultes, contrairement au cluster 2 avec 201 mots d'insultes et seulement 5,09 dans le cluster 1. Dans ce premier cluster regroupant les vidéos avec les espaces des commentaires particulièrement agressifs, trois médias — *Mediapart*, *Le Monde* et *Le HuffPost* — sont surreprésentés par rapport aux autres acteurs dans notre échantillon.

Graphique 1. Chaînes médias dans le cluster 1 : 55 vidéos avec l'agressivité la plus élevée dans les commentaires



- 50 Les différences entre les clusters sont moins remarquables lorsque l'on regarde la distribution des valeurs moyennes par commentaire.

Tableau 3. Valeurs moyennes des caractéristiques par commentaire par cluster

Cluster	N <sup>bre</sup> de mots	N <sup>bre</sup> de mots d'insulte	N <sup>bre</sup> de mots allongés	N <sup>bre</sup> de mots en majuscules	N <sup>bre</sup> d'exclamations	N <sup>bre</sup> d'interrogations
1	45,39	0,22	0,07	0,46	0,40	0,39
2	34,75	0,18	0,08	0,39	0,40	0,31
3	18,27	0,00	0,00	0,00	0,25	0,00

- 51 Le tableau 3 montre que si les trois clusters se distinguent considérablement en ce qui a trait à la longueur moyenne des commentaires (nombre de mots moyen par commentaire), les attributs d'agressivité sont assez proches, au moins pour les clusters 1 et 2. Ce résultat nous renseigne de deux façons. Premièrement, pour que les valeurs moyennes par commentaire soient proches entre les clusters, le nombre de commentaires doit être significativement plus important dans le cluster 1 (à forte concentration d'attributs d'agressivité). Deuxièmement, la concentration des attributs d'agressivité (clusters 1 et 2) est caractéristique des commentaires plus longs. Le tableau 4 confirme ces conclusions : en moyenne, par vidéo, le nombre de discussions ainsi que le nombre de commentaires sont beaucoup plus élevés dans les clusters 1 et 2 que dans le cluster 3 (respectivement 275,5, 92,15 et 3,5).

Tableau 4. Nombre de commentaires et de discussions dans les clusters

Cluster	Moyenne de n <sup>bre</sup> de discussions	Moyenne de n <sup>bre</sup> de commentaires
1	275,47	3 644,24
2	92,15	1 045,19
3	3,26	26,93
<b>Total</b>	<b>5,04</b>	<b>47,92</b>

- 52 Les interactions dans les espaces de commentaires des médias spécialisés semblent moins intenses que dans les espaces des médias généralistes ou partisans : le nombre de commentaires, le nombre de discussions et la longueur moyenne des commentaires sont plus faibles dans les médias spécialisés (tableau 5). Les catégories des médias ne sont pas représentées de la même manière dans les trois clusters des vidéos. Tandis que les médias généralistes sont présents dans tous les clusters, les médias spécialisés se retrouvent surtout dans le cluster 1, à faible niveau d'agressivité, et les partisans surtout dans le cluster 2, à forte agressivité (tableau 6).

Tableau 5. Interactions selon la catégorie des chaînes médias

	Moyenne de n <sup>bre</sup> de discussions	Moyenne de n <sup>bre</sup> de commentaires	Moyenne de n <sup>bre</sup> de mots par commentaire
<b>Médias généralistes</b>	5,3	49,3	25,3
<b>Partisans</b>	6,7	70,8	38,3
<b>Médias spécialisés</b>	2,6	24,5	16,8
<b>Total</b>	5,0	47,9	25,3

Tableau 6. Catégories des chaînes médias dans les clusters

	Agressivité faible (%)	Agressivité forte (%)	Agressivité moyenne (%)
<b>Médias généralistes</b>	79,1	60,0	809
<b>Partisans</b>	8,0	38,2	160
<b>Médias spécialisés</b>	12,9	1,8	30
<b>Total</b>	100,0	100,0	100,0

- 53 Une dernière observation intéressante (tableau 7) concerne la présence des commentateurs très populaires (avec un nombre de *likes* supérieur à la moyenne dans l'échantillon) et très actifs (qui commentent plus de trois chaînes différentes). En moyenne, par vidéo, le nombre de commentateurs populaires est particulièrement élevé dans les médias partisans, suivis par les généralistes (16,6 et 11,2 respectivement, la moyenne générale étant égale à 11). Par comparaison, les médias spécialisés n'en comptent en moyenne que 5,7. D'une part, cela signifie que les commentateurs les plus populaires ne sont pas forcément attirés par les médias spécialisés. D'autre part, les gestes d'appréciation et de distinction (comme les *likes*) sont probablement moins courants (ou moins importants compte tenu de l'homogénéité du public), les nombres de commentaires et de discussions étant plus faibles.
- 54 De même, les médias spécialisés tendent à moins attirer les commentateurs très actifs contribuant à plusieurs médias. En moyenne, seulement 6,3 commentateurs de ces chaînes commentent les vidéos de plus de 3 chaînes YouTube (la moyenne globale de toutes les vidéos dans la base étant égale à 14,5). Cet indicateur atteint son maximum pour les médias partisans (17,1 en moyenne) et les médias généralistes avec, en moyenne, 15,5 membres très actifs présents dans leurs vidéos.
- 55 D'un côté, ce résultat semble contre-intuitif compte tenu du fait que les contenus des médias spécialisés sont restreints par leurs orientations thématiques, ce qui devrait pousser leurs consommateurs à diversifier les sources d'information et à augmenter le nombre de chaînes consultées. D'un autre côté, il est possible que le public choisissant ces médias sur la plateforme YouTube ne recherche que des informations spécifiques et, de ce fait, ne commente que les chaînes diffusant le bon type de contenu. En ce qui concerne les médias partisans, les résultats confirment indirectement certaines observations des travaux précédents, présentées dans la revue de littérature : en particulier, les zones de discussions partisans peuvent attirer, en plus du public polarisé, politiquement et idéologiquement homogène, les utilisateurs d'autres médias ne partageant pas forcément les mêmes valeurs.

Tableau 7. Présence des commentateurs les plus populaires et les plus actifs dans les catégories des chaînes médias

	Moyenne des auteurs > 3 chaînes	Moyenne des auteurs populaires
Médias généralistes	15,5	11,2
Partisans	17,1	16,6
Médias spécialisés	6,3	5,7
Total	14,5	11,0

## Facteurs favorisant l'agressivité dans les espaces de commentaires

- 56 Avant de passer à l'analyse des facteurs à l'origine du débat agressif dans les espaces de commentaires sur YouTube et de vérifier nos hypothèses, nous construisons une nouvelle variable continue, le score d'agressivité, qui prend des valeurs de 0 à 1<sup>7</sup>.
- 57 Le tableau 8 présente les résultats de l'impact des différentes variables sur le score d'agressivité dans les espaces de commentaires des médias français (cf. annexe 2 pour plus de détails).

Tableau 8. Résultats des modèles de régressions

Score d'agressivité	Variables	Effet sur l'agressivité
Catégories	Médias spécialisés	<i>négatif</i> ***
	Médias généralistes	<i>positif</i> ***
	Médias partisans	<i>positif</i> ***
Espaces actifs	N <sup>bre</sup> de commentaires	<i>positif</i> ***
	N <sup>bre</sup> de commentateurs uniques	<i>positif</i> ***
	N <sup>bre</sup> de discussions	<i>positif</i> ***
Présence des membres « communautaires »	Proportion des auteurs avec plus de <i>likes</i>	<i>positif</i> ***
	Proportion des auteurs présents sur plus de 3 chaînes	<i>positif</i> ***

\*\*\* : coefficients très significatifs

- 58 Les variables fortement corrélées — nombre de discussions, nombre de commentaires et nombre de commentateurs uniques — ont été analysées séparément dans les modèles de régression. Pour l'ensemble des modèles obtenus, le pseudo-R<sup>2</sup> est supérieur à 0,2, ce qui indique un ajustement correct (McFadden, 1977).
- 59 Nos résultats montrent que, en lien avec la revue de la littérature et par opposition aux médias spécialisés, les chaînes des médias généralistes et les médias partisans abritent plus de discussions agressives. Ce résultat confirme notre hypothèse 1 et peut notamment s'expliquer par le fait que les médias spécialisés attirent un public plus homogène, échangeant autour de sujets d'intérêt commun pouvant échapper aux clivages partisans. Il est à ce titre possible pour les participants de faire émerger des règles de participation implicites sanctionnant les conduites socialement déviantes. Il faut toutefois noter qu'une partie importante de cette régulation échappe à notre

approche, ce résultat pouvant aussi être le produit des pratiques de modération de commentaires, en partie exercées par les propriétaires et animateurs des chaînes.

- 60 Au contraire, les médias généralistes attirent un public très hétérogène, et qui pour cette raison ne partage pas forcément de valeurs, de langage ou un même de vocabulaire. Plutôt qu'une action collective, la participation de ces commentateurs aux espaces de discussion de la plateforme peut être un moyen de manifester un point de vue, un désaccord avec les médias centraux ou une position personnelle face aux sujets d'actualité. Par conséquent, les espaces de débat des médias généralistes sont très marqués par l'agressivité.
- 61 Nous remarquons que, même si l'effet des médias partisans peut varier d'un à l'autre, ces espaces d'échanges sont généralement très marqués par l'agressivité. En effet, tandis que certains médias partisans ont des scores d'agressivité qui n'excèdent pas ceux des médias spécialisés (par exemple, *Lundi Matin*, *Acrimed* ou *Alternatives Économiques*), les autres semblent être affectés par une agressivité omniprésente.
- 62 Ensuite, nous observons que l'agressivité est plus présente dans les espaces de débat avec un nombre de commentaires et de commentateurs élevé, où les fils de discussion vont eux aussi être nombreux (hypothèse 2). En d'autres termes, le débat agressif apparaît davantage dans les espaces de commentaires plutôt développés, où les arguments sont nombreux, et non dans les espaces où les échanges sont moins fréquents et argumentés. En revanche, l'agressivité est faible dans les espaces de commentaires avec des taux d'interaction et de participation relativement faibles (notamment là où l'espace de commentaires est dominé par les commentaires *top-level* et où le nombre de vrais fils de discussion est faible) ; s'il n'y a pas d'échange, il n'y a pas de dispute.
- 63 Enfin, l'hypothèse 3 n'est pas validée par cette étude. Nous constatons que la proportion de commentateurs actifs et populaires a un impact positif et significatif sur le score d'agressivité. Plus le nombre de ces participants est élevé par rapport au nombre total de commentateurs, plus l'espace de commentaires est concerné par le problème de la conflictualité. Ce résultat pourrait être interprété de différentes manières. D'une part, sans forcément être à l'origine de la haine, les commentateurs « communautaires » de la plateforme peuvent être plus intéressés par les espaces plus polémiques et dynamiques des commentaires, espaces qui suscitent plus d'agressivité. D'autre part, leur présence dans les discussions sur les vidéos peut attirer l'attention des autres commentateurs. Ainsi, le nombre des participants augmentant, l'agressivité des débats s'en trouve favorisée, stimulée par les comportements des nouveaux entrants et par le passage à une activité nous ramenant aux observations de l'hypothèse 2 : plus il y a d'activité, plus il y a d'agressivité.
- 64 Une autre explication peut être liée à la dimension collective et communautaire des participants actifs et populaires. Dans le cadre du présent travail, nous ne cherchons pas à analyser les structures communautaires sous-jacentes qui lient ces membres entre eux. Or, il est possible que ces participants fassent partie de communautés différentes, opposant défenseurs de points de vue divergents. Dans ce cas, la participation simultanée des membres de ces communautés pourrait favoriser l'agressivité dans le débat. Comme cela a été proposé dans certains travaux (Pfeffer *et al.*, 2014), une analyse des communautés de la plateforme YouTube, du profil des participants et de leurs comportements serait une approche à privilégier dans les futures extensions de cette étude.

## Conclusion

- 65 Nos résultats montrent que le débat n'est pas nécessairement agressif ni insultant sur YouTube : un pourcentage non négligeable (20 %) des vidéos de notre étude n'a pas d'attributs d'agressivité dans les espaces de commentaires. A priori, malgré les avertissements retrouvés dans la littérature, le ton agressif et les débats brutaux ne sont pas forcément caractéristiques des espaces de commentaires dans la sphère francophone de YouTube. Ce résultat est valable pour toutes les catégories des médias : même les chaînes des médias généralistes (comme *Le HuffPost*), particulièrement touchées par l'agressivité, ont des vidéos (11,5 % pour *Le HuffPost*) où prennent place des discussions publiques non insultantes.
- 66 Néanmoins, la tendance globale est plus inquiétante. On constate que les espaces de commentaires très actifs, attirant plusieurs personnes différentes (par opposition aux cas où plusieurs commentaires sont le résultat de la participation de quelques commentateurs) et générant de vraies discussions (en plus des commentaires du type *top-level*), accueillent, en général, une forme de débat plus brutale. C'est peut-être pour cette raison que plusieurs médias agissent de manière radicale en verrouillant sur leurs chaînes YouTube les espaces de commentaires de certaines vidéos (par exemple, *Le Télégramme*). Ce problème peut aussi être à l'origine du fait que, pour plusieurs vidéos des chaînes YouTube des médias français étudiés, les utilisateurs préfèrent de ne pas laisser de commentaire, tout en ignorant ou en négligeant la possibilité de participer à une discussion publique offerte par la plateforme. En effet, après la collecte initiale des données sur YouTube, nous avons récupéré les informations pour plus de 137 K de vidéos. Cependant, seulement 46 K (33,5 %) de ces vidéos ont fait partie de cette étude, car les autres ne contiennent aucun commentaire.
- 67 Nos résultats montrent également qu'un moyen efficace pour diminuer l'agressivité du débat et pour favoriser l'engagement « positif » des utilisateurs dans la communication médiatisée est de minimiser l'anonymat et d'attirer dans ces discussions un public ayant en commun des centres d'intérêt. Cet objectif peut être atteint grâce à un positionnement explicite d'un média. C'est clairement le cas des médias spécialisés, mais aussi de certains médias généralistes, comme *Ouest-France* ou *La Croix* qui comptent des pourcentages relativement importants de vidéos aux espaces de commentaires dénués d'agressivité (44,6 % et 33,5 % respectivement). Suivant cette logique, la non-agressivité pourrait aussi être stimulée par la plateforme elle-même au moyen de ses dispositifs d'interaction entre les utilisateurs (par exemple, la possibilité de constituer son réseau de contacts, de créer un profil utilisateur détaillé, de partager certains contenus dans un groupe).
- 68 Pour les médias partisans, nos résultats confirment certaines observations des recherches précédentes et contribuent à l'analyse du lien entre la polarisation et l'agressivité. On observe notamment que, malgré une homogénéité générale du public dans ces espaces de discussion, les orientations explicitement politiques favorisent l'occurrence de prises de parole contestataires, et parfois sciemment outrancières, ce qui a souvent pour résultat de faire glisser la discussion vers un affrontement verbal agressif. L'analyse de cette question restant sous-développée dans les travaux existants, elle mériterait d'être explorée plus profondément (Humprecht *et al.*, 2020).

69 Finalement, cet article met en évidence que certains médias, par leur image ou leur statut de média de référence, attirent un public hétérogène et parfois brutal. C'est par exemple le cas du *Monde* ou *Mediapart* et *AgoraVox* chez les médias partisans (qui sont aujourd'hui les médias de référence dans le domaine de la presse citoyenne), médias cristallisant une certaine défiance envers le journalisme politique. Un autre média partisan, *Fdesouche*, affiche aussi des scores d'agressivité particulièrement élevés et stables. *Fdesouche*, qui se présente comme un agrégateur d'informations (très sélectionnées et orientées), est également un média de référence très influent de l'extrême droite (Albertini et Doucet, 2016). En plus, dans ce cas, l'anonymat (favorisant, comme on l'a vu plus haut, l'agressivité) est au cœur du fonctionnement même du média : certains médias partisans de notre échantillon (*Fdesouche*, *Lundi Matin* et *Contre-info*) ne publient aucune information sur leurs rédactions, ont des membres qui restent anonymes, ont la plupart des articles publiés sans signature des auteurs et proposent des sites qui ne fournissent pas d'informations biographiques sur leurs contributeurs. Ces exemples montrent que la présence d'un débat agressif et insultant n'est pas forcément liée à la plateforme YouTube elle-même et nous invitent à considérer l'analyse comparative d'autres plateformes numériques comme nécessaire pour mieux comprendre ce phénomène.

WRIGHT Scott, Todd GRAHAM et Dan JACKSON (2016), « Third space, social media, and everyday political talk », dans Axel BRUNS, Gunn ENLI, Eli SKOGERBO, Anders OLOF et Christian CHRISTENSEN (dir.), *The Routledge Companion to Social Media and Politics*, New York, Routledge, p. 74-88.

---

## BIBLIOGRAPHIE

- ADOMAVICIUS, Gediminas, TUZHILIN, Alexander (2005), « Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions », *IEEE Transactions on Knowledge and Data Engineering*, 17(6), p. 734-749. <https://ieeexplore.ieee.org/abstract/document/1423975>. Page consultée le 6 mai 2021.
- AGRESTI Alan (2018), *An Introduction to Categorical Data Analysis*, Hoboken, Ed. John Wiley & Sons.
- AKAIKE Hirotogu (1998), « Information theory and an extension of the maximum likelihood principle », dans *Selected Papers of Hirotogu Akaike*, New York, Springer, p. 199-213, [https://link.springer.com/chapter/10.1007/978-1-4612-1694-0\\_15](https://link.springer.com/chapter/10.1007/978-1-4612-1694-0_15), page consultée le 6 mai 2021.
- ALBERTINI Dominique et David DOUCET (2016), *La Fachosphère. Comment l'extrême droite remporte la bataille d'Internet*, Paris, Flammarion.
- AMOSSY Ruth et Marcel BURGER (2011), « Introduction : la polémique médiatisée », *SEMEN – Revue de sémio-linguistique des textes et discours*, 31, <https://journals.openedition.org/semen/9072>, page consultée le 6 mai 2021.
- ANDERSON Ashley A., Sara K. YEO, Dominique BROSSARD, Dietram A. SCHEUFELE et Michael A. XENOS (2016), « Toxic talk: How online incivility can undermine perceptions of media »,



- COVINGTON Paul, Jay ADAMS et Emre SARGIN (2016), « Deep neural networks for YouTube recommendations », *Proceedings of the 10th ACM Conference on Recommender Systems*, New York, Association for Computing Machinery, p. 191-198.
- DAVIS Richard (1999), *The Web of Politics: The Internet's Impact on the American Political System*, Oxford, Oxford University Press
- DELFORCE Bernard (2010), « Discursivité sociale / discours sociaux : penser les enjeux sociaux de l'information », dans Aurélie TAVERNIER et al. (dir.), *Figures sociales des discours. Le « discours social » en perspectives*, Lille, Édition du Conseil scientifique de l'Université Charles-de-Gaulle, p. 23-42.
- EDGERLY Stephanie, Emily K. VRAGA, Timothy K.F. FUNG et Tae JOON (2008), « Youtube as a public sphere: The Proposition 8 debate », *Association of Internet Researchers conference*, Milwaukee, WI.
- ESPINHEIRA Patricia, Silvia FERRARI et Francisco CRIBARI-NETO (2008), « On beta regression residuals », *Journal of Applied Statistics*, 35(4), p. 407-419.
- GARRETT Kelly, R. (2009), « Echo chambers online? Politically motivated selective exposure among internet users », *Journal of Computer Mediated Communication*, 14(2), p. 265-285, <https://academic.oup.com/jcmc/article/14/2/265/4582957?login=true>, page consultée le 6 mai 2021.
- GEORGES Fanny (2019), « La chasse aux fantômes sur YouTube. Approche ethnographique et quali-quantitative des commentaires des vidéos », *Revue de communication sociale et publique*, 27, p. 99-122, <https://journals.openedition.org/communiquer/4989>, page consultée le 6 mai 2021.
- GRAHAM Todd (2012), « Beyond “political” communicative spaces: Talking politics on the wife swap discussion forum », *Journal of Information Technology and Politics*, 9(1), p. 31-45.
- GUIBERT Gérard, Franck REBILLARD et Fabrice ROCHELANDET (2016), *Médias, culture et numérique : approches socioéconomiques*, Paris, Armand Colin.
- HAIR Joseph F., Rolph E. ANDERSON, Ronald L. TATHAM et William C. BLACK (1995) « Multivariate data analysis », dans John A. HARTIGAN et Manchek A. WONG (1979), « Algorithm AS 136: A k-means clustering algorithm », *Applied Statistics*, 28, p. 100-108.
- HALPERN Daniel et Jennifer GIBBS (2013), « Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression », *Computers in Human Behavior*, 29, p. 1159-1168.
- HEMALATHA Indukuri, Gottumukala P. VARMA, G.P. et Saradhi A GOYARDHAN (2012), « Preprocessing the informal text for efficient sentiment analysis », *International Journal of Emerging Trends & Technology in Computer Science*, 1(2), p. 58-61, [https://scholar.google.ca/scholar?hl=fr&as\\_sdt=0%2C5&q=Preprocessing+the+informal+text+for+efficient+sentiment+analysis&btnG=](https://scholar.google.ca/scholar?hl=fr&as_sdt=0%2C5&q=Preprocessing+the+informal+text+for+efficient+sentiment+analysis&btnG=), page consultée le 6 mai 2021.
- HSUEH Mark, Kuma YOGESWARAN et Sanna MALINEN (2015), « “Leave your comment below”: Can biased online comments influence our own prejudicial attitudes and behaviors? », *Human Communication Research*, 41(4), p. 557-576, <https://academic.oup.com/hcr/article/41/4/557/4064434?login=true>, page consultée le 6 mai 2021.
- HUMPRECHT Edda et Frank ESSER (2018), « Diversity in online news », *Journalism Studies*, 19(12), p. 1825-1847.
- HUMPRECHT Edda, Lea HELLMUELLER et Juliane LISCHKA (2020), « Hostile emotions in news comments: A cross-national analysis of Facebook discussions », *Social Media + Society*, 6(1), <https://journals.sagepub.com/doi/full/10.1177/2056305120912481>, page consultée le 6 mai 2021.

- HYNDMAN Rob J. et Anne KOEHLER (2006), « Another look at measures of forecast accuracy », *International Journal of Forecasting*, 22(4), p. 679-688.
- JANE Emma (2015), « Flaming? What flaming? The pitfalls and potentials of researching online hostility », *Ethics and Information Technology*, 17(1), p. 65-87.
- KHAN Laeeq et Solomon JACOB (2013), « Advocacy, entertainment and news — An analysis of user participation on Youtube », SSRN, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2579596](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2579596), page consultée le 23 mai 2021.
- KWON K. Hazel et Daegon CHO (2017), « Swearing effects on citizen-to-citizen commenting online: A large-scale exploration of political versus nonpolitical online news sites », *Social Science Computer Review*, 35(1), p. 84-102.
- KWON K. Hazel et Anatoliy GRUZD (2017), « Is aggression contagious online? A case of swearing on Donald Trump's campaign videos on Youtube », *Actes du colloque 50th Hawaii International Conference on System Sciences*, Waikoloa Village, IEEE.
- LADHARI Riadh, Elodie MASSA et Hamida SKANDRANI (2020), « Youtube vloggers' popularity and influence: the roles of homophily, emotional attachment, and expertise », *Journal of Retailing and Consumer Services*, 54.
- LAINÉ Michael S.S., Gunes ERCAL et Bo LUO (2011), « User groups in social networks: an experimental study on Youtube », *Actes du colloque 44th Hawaii International Conference on System Sciences*, IEEE, <https://ieeexplore.ieee.org/abstract/document/5718493>, page consultée le 6 mai 2021.
- LEWIS David D. (1992), « Feature selection and feature extraction for text categorization », *Actes du colloque Workshop on Speech and Natural Language*, Association for Computational Linguistics, p. 212-217, [https://scholar.google.ca/scholar?hl=fr&as\\_sdt=0%2C5&q=Feature+selection+and+feature+extraction+for+text+categorization+&btnG=](https://scholar.google.ca/scholar?hl=fr&as_sdt=0%2C5&q=Feature+selection+and+feature+extraction+for+text+categorization+&btnG=), page consultée le 6 mai 2021.
- LOPS Pasquale, Marco DE GEMMIS, Giovanni SEMERARO, Cataldo MUSTO, Fedelucio NARDUCCI et Massimo BUX (2011), « A semantic content-based recommender system integrating folksonomies for personalized access », dans Giovanna CASTELLANO, Jain LKHMI et Anna Maria FANELLI (dir.), *Web Personalization in Intelligent Environment*, Berlin, Springer, p. 73-105.
- LUXTON David D., Jennifer D. JUNE et Jonathan M. FAIRALL (2012), « Social media and suicide: A public health perspective », *American Journal of Public Health*, 102(2), p. 195-S200.
- LYUBAREVA Inna, Fabrice ROCHELANDET et Yanis HARALAMBOUS (2020), « Qualité et différenciation des biens informationnels. Une étude exploratoire sur l'information d'actualité », *Revue d'économie industrielle*, 172, p. 133-177.
- LYUBAREVA Inna et Fabrice ROCHELANDET (2016), « L'évolution des modèles d'affaires dans les industries créatives : l'exemple de la presse en ligne en France (2004-2014) », *Revue d'économie industrielle*, 156, p. 123-157, <https://journals.openedition.org/rei/6460>, page consultée le 6 mai 2021.
- LYUBAREVA Inna, Laurent BRISSON, Cécile BOTHOREL et Romain BILLOT (2020), « Une plateforme de crowdfunding et son réseau social. L'exemple Ulule », *Revue française de gestion*, 286, p. 135-151.
- MARQUES José, Dominic ABRAMS et Rui G. SERÔDIO (2001), « Being better by being right: Subjective group dynamics and derogation of in-group deviants when generic norms are undermined », *Journal of Personality and Social Psychology*, 81(3), p. 436-447.

- MCFADDEN Daniel (1977), *Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments*, Berkeley, Institute of Transportation Studies, University of California.
- MCPHERSON Miller, Lynn SMITH-LOVIN et James COOK (2001), « Birds of a feather: Homophily in social networks », *Annual Review of Sociology*, 27(1), p. 415-444.
- MESANGEAU Julien et Céline MORIN (à paraître), « La liminalité d'un contre-public sur YouTube : étude des rituels d'intégration en ligne d'un contre-public hors-ligne », *Terminal*.
- MILLER Patrick R. et Pamela Johnston CONOVER (2015), « Red and blue states of mind: Partisan hostility and voting in the United States », *Political Research Quarterly*, 68(2), p. 225-239.
- MOOR Peter J., Ard HEUVELMAN et Ria VERLEUR (2010), « Flaming on Youtube », *Computers in Human Behavior*, 26(6), p. 1536-1546.
- MUDDIMAN Ashley et Natalie Jomini STROUD (2017), « News values, cognitive biases, and partisan incivility in comment sections », *Journal of Communication*, 67, p. 586-609, <https://academic.oup.com/joc/article/67/4/586/4642138>, page consultée le 6 mai.
- NETO Joaquim, Yokoyama M. KAZUKI et Karin BECKER (2017) « Studying toxic behavior influence and player chat in an online video game », Actes du colloque *International Conference on Web Intelligence*, 26-23, Leipzig, Germany.
- NOBATA Chikashi, Joel TETREAU, Thomas ACHINT *et al.* (2016), « Abusive language detection in online user content », *Proceedings of the 25th International Conference on World Wide Web*, Genève, International World Wide Web Conferences Steering Committee, p. 145-153.
- PEIS MORALES DEL CASTILLO José et Ja DELGADO LOPEZ (2008), « Semantic recommender systems: analysis of the state of the topic », *Hipertext.net*, 6, [https://scholar.google.com/scholar?hl=fr&as\\_sdt=0%2C5&q=Semantic+recommander+systems%3A+analysis+of+the+state+of+the+topic&btnG=](https://scholar.google.com/scholar?hl=fr&as_sdt=0%2C5&q=Semantic+recommander+systems%3A+analysis+of+the+state+of+the+topic&btnG=), page consultée le 6 mai 2021.
- PFEFFER Jürgen, Thomas ZORBACH, Kathleen M. CARLEY (2014), « Understanding online firestorms: Negative word-of-mouth dynamics in social media networks », *Journal of Marketing Communications*, 20(1-2), p. 117-128.
- PRIMACK Brian A., Ariel SHENSA, Jaime E. SIDANI, Erin O. WHAITE *et al.* (2017), « Social media use and perceived social isolation among young adults in the US », *American Journal of Preventive Medicine*, 53(1), p. 1-8.
- RAPPORT DE L'INSTITUT MONTAIGNE (2019), *Unfolding the multi-layered structure of the French mediascape*, préparé par Dominique CARDON, Jean-Philippe COINTET, Benjamin OOGHE et Guillaume PLIQUE, Paris, Sciences Po Médialab.
- REBILLARD Franck et Nikos SMYRNAIOS (2010), « Les infomédiaires, au cœur de la filière de l'information en ligne », *Réseaux*, 160-161, p. 163-194, <https://www.cairn.info/revue-reseaux-2010-2-page-163.htm>, page consultée le 6 mai 2021.
- REMY Lionel (2018), « Entre islam et théories du complot, idéations religieuses et identités sur YouTube », *Recherches sociologiques et anthropologiques*, 49(1), p. 49-78, <https://journals.openedition.org/rso/2391>, page consultée le 6 mai 2021.
- RHEINGLOD Howard (2000), *The Virtual Community: Homesteading on the Electronic Frontier*, Cambridge, MIT press.
- RIPLEY Brian et William VENABLES (2016), « "Package "nnet" », *R package version 7*, p. 3-12, <https://cran.r-project.org/web/packages/nnet/nnet.pdf>, page consultée le 24 mai 2021.

- RÖSNER Leonie, Stephan WINTER et Nicole C. KRÄMER (2016), « Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior », *Computers in Human Behavior*, 58, p. 461-470.
- ROTMAN Dana et Jennifer PREECE (2010), « The “WeTube” in YouTube — Creating an online community through video sharing », *International Journal of Web Based Communities*, 6(3), p. 317-333.
- SCHACHTER Stanley (1951), « Deviation, rejection, and communication », *The Journal of Abnormal and Social Psychology*, 46(2), p. 190-207.
- SCHULTES Peter, Verena DORNER et Franz LEHNER (2013), « Leave a comment! An in-depth analysis of user comments on Youtube », *Wirtschaftsinformatik*, 42, p. 659-673, [https://scholar.google.com/scholar?hl=fr&as\\_sdt=0%2C5&q=Leave+a+comment%21+An+in-depth+analysis+of+user+comments+on+Youtube&btnG=](https://scholar.google.com/scholar?hl=fr&as_sdt=0%2C5&q=Leave+a+comment%21+An+in-depth+analysis+of+user+comments+on+Youtube&btnG=), page consultée le 6 mai 2021.
- SCOTT LONG John (1997), « Regression models for categorical and limited dependent variables », *Série Advanced Quantitative Techniques in the Social Sciences*, 7, Thousand Oaks, Sage.
- SULER John (2004), « The online disinhibition effect », *Cyberpsychology & Behavior*, 7(3), p. 321-326.
- SUNSTEIN Cass (2001), *Republic.com*, Princeton, Princeton University Press.
- SUNSTEIN Cass (2018), *#Republic: Divided Democracy in the Age of Social Media*, Princeton, Princeton University Press.
- TUOMARLA Ulla (2014), « La recontextualisation et la circulation d’insultes dans les médias: le cas des commentaires des lecteurs sur des articles/vidéos publiés en ligne », *Cahiers de praxématique*, 63, <https://journals.openedition.org/praxematique/2420>, page consultée le 6 mai 2021.
- VINCENT Diane et Geneviève BERNARD BARBEAU (2012), « Insulte, disqualification, persuasion et tropes communicationnels ; à qui l’insulte profite-t-elle ? », *Argumentation et analyse du discours*, 8, <https://journals.openedition.org/aad/1252>, page consultée le 6 mai 2021.
- VOGEL Erin A., Jason P. ROSE, Lindsay R. ROBERTS et Katheryn ECKLES (2014), « Social comparison, social media, and self-esteem », *Psychology of Popular Media Culture*, 3(4), p. 206-222.
- VON HIPPEL Eric (2005), *Democratizing Innovation*, Cambridge, MIT press.
- WACHS Sebastian et Michelle F. WRIGHT (2018), « Associations between bystanders and perpetrators of online hate: The moderating role of toxic online disinhibition », *International Journal of Environmental Research and Public Health*, 15(9), p. 2030.
- WEBER Patrick (2014), « Discussions in the comments section: Factors influencing participation and interactivity in online newspapers’ reader comments », *New Media & Society*, 16(6), p. 941-957.
- WOJCIESZAK Magdalena E. et Diana C. MUTZ (2009), « Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? », *Journal of Communication*, 59, p. 40-56. <https://academic.oup.com/joc/article/59/1/40/4098524?login=true>, page consultée le 6 mai 2021.
- WRIGHT Scott et John STREET (2007), « Democracy, deliberation and design: The case of 31 online discussion forums », *New Media & Society*, 9(5), p. 849-869.
- WRIGHT Scott (2016), « Les conversations politiques en ligne au quotidien : design, délibération et “tiers espaces” », *Questions de communication*, 30, p. 119-134. <https://journals.openedition.org/questionsdecommunication/10734>, page consultée le 6 mai 2021.

ZEILEIS Achim et Francisco CRIBARI-NETO (2010), « Beta regression in R », *Journal of Statistical Software*, 34(2), p. 1-24.

ZIEGELE Marc, Mathias WEBER et Oliver QUIRING (2018), « The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions », *Information, Communication & Society*, 21, p. 1419-1435.

## ANNEXES

Tableau 9. Liste des chaînes et nombre de leurs vidéos dans l'échantillon

Généralistes	N <sup>bre</sup> de vidéos	Spécialisées	N <sup>bre</sup> de vidéos
<i>20 Minutes FR</i>	588	<i>Corse Net Infos</i>	111
<i>Groupe Nice-Matin</i>	422	<i>Corse-Matin Presse</i>	107
<i>France Net Infos</i>	27	<i>Gala</i>	170
<i>CNEWS</i>	8 950	<i>ZDNet.fr</i>	2
<i>Courrier international</i>	81	<i>La Manche Libre</i>	55
<i>Franceinfo</i>	3 671	<i>Orient XXI</i>	39
<i>France-Soir</i>	432	<i>L'Équipe</i>	4 855
<i>La Croix</i>	453	<i>StreetPress</i>	308
<i>LCI</i>	210	<i>Reporterre</i>	66
<i>La Voix du Nord</i>	34	<i>Hors-Série</i>	69
<i>Le Télégramme</i>	216	<i>Famille Chrétienne</i>	132
<i>La Provence</i>	674	<b>Total</b>	<b>5 914</b>
<i>Le Figaro</i>	4 724		
<i>Le Monde</i>	1 385	<b>Partisanes</b>	<b>N<sup>bre</sup> de vidéos</b>
<i>Le Parisien</i>	1 834	<i>Acrimed</i>	55
<i>Le Point</i>	604	<i>AgoraVox</i>	179
<i>Le HuffPost</i>	5 141	<i>Alternatives Économiques</i>	31
<i>Les Échos</i>	841	<i>Causeur</i>	86
<i>L'Express</i>	819	<i>Contre-info</i>	23
<i>Libération</i>	12	<i>Fakirpresse</i>	77
<i>L'Obs</i>	1 854	<i>Fdesouche</i>	7
<i>Metronews</i>	30	<i>Là — bas si j'y suis</i>	286
<i>Ouest-France</i>	994	<i>Le Monde diplomatique</i>	7
<i>Paris Match</i>	1 146	<i>Lundi Matin</i>	89
<i>Rue89</i>	68	<i>Lutte ouvrière</i>	30
<i>Slate.fr</i>	102	<i>Valeurs Actuelles</i>	111
<i>Sud Ouest</i>	601	<i>Atlantico</i>	1
<i>La Tribune</i>	328	<i>L'Humanité</i>	799
<i>Le Dauphiné Libéré</i>	219	<i>Arrêt sur Images</i>	85
<i>Dernières Nouvelles d'Alsace</i>	10	<i>Mediapart</i>	1 840
<b>Total</b>	<b>36 470</b>	<b>Total</b>	<b>3 706</b>

Tableau 10. Premier modèle

Score agressivité	(1)	(2)
partisans	0,206*** (0,008)	0,493*** (0,017)
généralistes		0,287*** (0,016)
spécialisés	-0,287*** (0,016)	
message_count	0,0001*** (0,00000)	0,0001*** (0,00000)
prop_authors_like	2,180*** (0,025)	2,180*** (0,025)
prop_authors_more_3	0,283*** (0,016)	0,283*** (0,016)
Constant	-5,520*** (0,013)	-5,810*** (0,019)
Observations	46 088	46 088
pseudo-R <sup>2</sup>	0,232	0,232
Log Likelihood	167,562,000	167,562,000
<i>Note:</i> * p<0,1 ** p<0,05 *** p<0,01		

Tableau 11. Deuxième modèle

Score agressivité	(3)	(4)
généralistes		0,273*** (0,016)
spécialisées	-0,273*** (0,016)	
partisanes	0,213*** (0,008)	0,486*** (0,017)
discussion_count	0,001*** (0,00002)	0,001*** (0,00002)
prop_authors_like	2,130*** (0,026)	2,130*** (0,026)
prop_authors_more_3	0,261*** (0,016)	0,261*** (0,016)
Constant	-5,490*** (0,013)	-5,770*** (0,019)
Observations	46,088	46,088
Pseudo-R <sup>2</sup>	0,225	0,225
Log Likelihood	167,322,000	167,322,000
<i>Note:</i> * p<0,1 ** p<0,05 *** p<0,01		

Tableau 12. Troisième modèle

Score agressivité	(5)	(6)
généralistes		0,292***
		(0,016)
spécialisés	-0,292***	
	(0,016)	
partisans	0,206***	0,497***
	(0,008)	(0,017)
distinct_authors_count	0,0001***	0,0001***
	(0,00000)	(0,00000)
prop_authors_like	2,230***	2,230***
	(0,025)	(0,025)
prop_authors_more_3	0,285***	0,285***
	(0,016)	(0,016)
Constant	-5,520***	-5,820***
	(0,013)	(0,019)
Observations	46,088	46,088
pseudo-R <sup>2</sup>	0,231	0,231
Log Likelihood	167,458,000	167,458,000
Note: *p<0,1 **p<0,05 ***p<001		

## NOTES

1. Nous distinguons dans l'analyse les commentaires *top-level* – les messages publiés après une vidéo, qui sont restés sans réaction sous forme de réponse directe à ces commentaires de la part des autres commentateurs – des commentaires dans les fils de discussion, qui sont les réponses aux commentaires *top-level*.

2. Cf. par exemple, [fr.wiktionary.org/w/index.php?title=Catégorie:Insultes\\_en\\_français&pageuntil=baleine#mw-pages](https://fr.wiktionary.org/w/index.php?title=Catégorie:Insultes_en_français&pageuntil=baleine#mw-pages), ou [strategicalblog.com/liste-dinsultes-francaises-pas-trop-vulgaires/](https://strategicalblog.com/liste-dinsultes-francaises-pas-trop-vulgaires/). Page consultée le 6 mai 2021. À ces listes nous avons ajouté d'autres lexiques que nous avons repérés lors de l'analyse manuelle.

3. Nous proposons ici quelques exemples de commentaires issus des discussions des chaînes de notre échantillon. Toutes les citations gardent leur orthographe originale.

4. Cette variable nous permet de tenir compte de vraies discussions, qui apparaissent comme réaction des utilisateurs à un commentaire publié sous forme de réponse à ce commentaire.

5. Étant donné que les proportions des vidéos dans les clusters générés ont été très déséquilibrées, l'algorithme *SMOTE* (Chawla *et al.*, 2002) a été appliqué afin de corriger ce travers et d'éviter le biais de sur-échantillonnage. *Synthetic Minority Over-Sampling Technique (SMOTE)* est une méthode utilisée souvent à l'étape de préparation des données. Elle consiste tout simplement à générer des observations synthétiques à partir de la base d'origine afin d'obtenir une représentativité équilibrée des différentes classes.

6. La qualité d'ajustement du modèle aux données a été validée par les méthodes classiques : pseudo-R<sup>2</sup> (Scott Long, 1997) ; la qualité statistique globale du modèle a été déterminée par le critère d'information d'Akaike (AIC, *Akaike information criterion* [Akaike, 1998]) ; et la puissance

prédictive des modèles a été évaluée par l'écart quadratique moyen (RMSE, *root-mean-square error* [Hyndman et Koehler, 2006]).

7. En prenant en compte les résultats présentés ci-dessus et, en particulier, le fort déséquilibre entre les clusters de vidéos de notre échantillon, nous appliquons la méthode *SMOTE* pour obtenir une distribution de données plus homogène. Ces transformations de données ainsi que l'attribution arbitraire des poids dans le calcul du *score* ont été précédées par une série de tests. Elles ont donc été retenues pour l'analyse finale à la suite de l'étude et de la comparaison de la performance des différents modèles (par exemple, pseudo-R<sup>2</sup>). Néanmoins, les résultats, avec et sans les traitements supplémentaires, ont été similaires, ce qui confirme la robustesse de l'analyse effectuée.

---

## RÉSUMÉS

Les auteurs cherchent à comprendre, d'une part, où et quand l'espace des commentaires sur la plateforme YouTube en vient à être touché par les débats brutaux ou agressifs et, d'autre part, quels facteurs limitent ou favorisent cette agressivité. Pour atteindre cet objectif, ils emploient un dispositif se positionnant à la lisière du traitement automatisé du langage et de l'économie politique des médias. Ils basent leur étude sur 2 209 206 commentaires, dont 1 184 859 se trouvent dans les fils de discussion, répartis dans les espaces de commentaires de 46 090 vidéos. Ces dernières sont issues d'un panel de 57 chaînes de médias français aux catégories institutionnelles et aux positionnements différents.

The authors seek to understand where and when YouTube comment sections are hit by brutal or aggressive debates and the factors that limit or foster this aggressive behaviour. To do so, they use a device at the crossroads of automated language processing and the political economy of media. The study is based on 2,209,206 comments, 1,184,859 of which are found in discussion threads, from the comment sections of 46,090 videos. The videos are taken from the YouTube channels of 57 different French media channels representing a variety of positions and institutional categories.

Los autores del artículo procuran comprender, por una parte, dónde y cuándo el espacio de los comentarios en las plataformas YouTube llega a ser afectado por debates violentos o agresivos y por otra parte, qué factores limitan o favorecen esta agresividad. Para lograr el objetivo, los autores del artículo emplean un dispositivo que se posiciona al límite del tratamiento automatizado del lenguaje y la economía política de los medios de comunicación. Dichos autores basan su estudio en 2 209 206 comentarios, entre los cuales, 1 184 859 se encuentran en los hilos de discusión, repartidos en los espacios comentarios de 46 090 vídeos. Estos últimos provienen de un panel de 57 canales de los medios de comunicación francesa en las categorías institucionales y posicionamientos diferentes.

## INDEX

**Palabras claves** : comportamiento agresivo, debates violentos, videos, YouTube canales de los medios, Francia

**Keywords** : aggressive behaviour, brutal debates, videos, YouTube media channels, France

**Mots-clés** : comportements agressifs, débats brutaux, vidéos, chaînes YouTube médias, France

## AUTEURS

### INNA LYUBAREVA

Inna Lyubareva est maître de conférences, IMT Atlantique et membre du Laboratoire d'économie et de gestion de l'Ouest (LEGO). Courriel : inna.lyubareva@imt-atlantique.fr

### JULIEN MESANGEAU

Julien Mesangeau est maître de conférences, Université Sorbonne Nouvelle-Paris 3 et membre de l'Institut de recherche médias, cultures, communication et numérique (Irméccen). Courriel : mesangeaujulien@gmail.com

### NADIRA BOUDJANI

Nadira Boudjani est postdoctorant, IMT Atlantique et membre du Laboratoire des sciences et techniques de l'information, de la communication et de la connaissance (Lab-STICC). Courriel : boudjaninadira@gmail.com

### IMAD EL BADISY

Imad El Badisy est ingénieur, IMT Atlantique et membre du Laboratoire d'économie et de gestion de l'Ouest (LEGO). Courriel : imad.el-badisy@imt-atlantique.fr

### LAURENT BRISSON

Laurent Brisson est maître de conférences, IMT Atlantique et membre du membre du Laboratoire des sciences et techniques de l'information, de la communication et de la connaissance (Lab-STICC). Courriel : laurent.brisson@imt-atlantique.fr