

Étude quantitative de l'intensité médiatique des six premiers mois de la pandémie du Covid-19

Nicolas Hervé, Institut National de l'Audiovisuel

RÉSUMÉ

Le but de cette étude est d'observer l'ampleur de la médiatisation du coronavirus sur différents supports (AFP, presse en ligne, chaînes télévisées et radio d'information en continu et Twitter), de les comparer, de les analyser et de les mettre en relation avec les événements clés de la chronologie de cette pandémie. Nous avons mis en place une approche quantitative sur un corpus complet couvrant les six premiers mois du Covid-19 dans l'espace médiatique français. Avec une approche textométrique incluant le calcul du temps d'antenne pour les médias audiovisuels, nous avons estimé la proportion de la production de contenu de tous les supports qui est dédiée à la pandémie. De plus, nous réalisons un focus particulier sur la médiatisation de Didier Raoult ainsi que sur l'impact de la pandémie sur le travail des journalistes dans les rédactions de presse en ligne.

ABSTRACT

The aim of this study is to observe the extent of media coverage of the coronavirus in different media (AFP news agency, online press, TV and radio news channels and Twitter). We compare them with each other, we analyze them and we relate them to key events in the chronology of the pandemic. We applied a quantitative approach to a complete corpus covering the first six months of the Covid-19 pandemic in the French media space. With a textometric approach including the calculation of airtime for audiovisual media, we estimated the proportion of content production of all media that is dedicated to the pandemic. In addition, we focus on the media coverage of Didier Raoult and the impact of the pandemic on the work of journalists in online newsrooms.

La pandémie du coronavirus est encore loin d'être finie, mais on peut déjà dire que sa médiatisation a eu un caractère exceptionnel. Quel que soit le canal considéré (télévision, radio, agences de presse, presse en ligne), l'ampleur et la durée du traitement médiatique qui lui ont été consacrés sont inédites dans l'histoire récente. La qualité intrinsèque de ce phénomène mondial explique probablement largement cette médiatisation. En effet, contrairement à certains événements tragiques (tels que les attentats) qui se sont déroulés en France mais sur des périodes très courtes ou à l'inverse des guerres et révoltes sur des durées plus longues mais à l'étranger et perçues comme lointaines, la pandémie cumule les deux critères de localisation sur tout le territoire et de crise sans fin. De plus, puisque l'ensemble de la population et des activités sociales et économiques a été touché par les conséquences directes ainsi que par les mesures mises en place, les rédactions elles-mêmes ont été concernées et les journalistes ont dû s'adapter.

Notre objectif est de décrire les six premiers mois de l'année 2020 dans les médias en France d'un point de vue extérieur, en analysant leur production journalistique par une approche quantitative. Nous proposons ainsi une chronologie retraçant pour chaque média la proportion de sa diffusion qui est consacrée à la pandémie. On peut ainsi observer les grandes phases de la médiatisation et détecter les changements dans les choix éditoriaux. Cela permet de comparer les médias et de dégager les grandes tendances. De plus, en rapprochant ces données des faits, on peut identifier les moments clés concomitants à ces changements (on parle bien ici de corrélation et non de causalité). Pour compléter ce tableau global, nous analysons également les propos échangés sur Twitter pendant la même période afin d'avoir un point de comparaison sur les sujets de discussion et leur éventuelle synchronisation avec l'agenda médiatique.

La spécificité de notre approche réside dans la complétude du corpus que nous utilisons ainsi que dans la méthode qui a été élaborée pour sa constitution. La quasi-exhaustivité des sources et la profondeur temporelle nous permettent d'éviter les biais classiques des analyses de corpus. La captation des documents a commencé bien avant le début de la pandémie et, surtout, n'est aucunement liée à la thématique des études que nous sommes amenés à faire puisqu'elle n'est pas basée sur des mots clés ou requêtes déterminés de façon ad hoc. Le corpus extrait pour cette étude s'étend de début décembre 2019 (période de contrôle hors pandémie) à fin juin 2020. Il contient l'intégralité des dépêches AFP, des publications en ligne de 38 sites d'information et environ 60 % des tweets émis en français. Pour la télévision et la radio, les contenus sont transcrits automatiquement et nous disposons donc sous forme textuelle de tout ce qui est prononcé à l'antenne. Nous avons choisi de transcrire les chaînes et radios d'information en continu entre 6 h et minuit. Disposant d'un corpus complet, nous pouvons alors raisonner en proportion de la production totale et ainsi comparer la presse en ligne et la télévision ou Twitter et la radio. Notre analyse est basée sur de la textométrie, complétée par une modélisation du temps d'antenne pour la télévision et la radio.

État de l'art : autres études

L'étude de corpus textuels a une longue tradition en France. La médiatisation des événements et leur visibilité dans la sphère publique en sont une part importante et font l'objet de discussions méthodologiques. Dalibert (2018) souligne ainsi que l'analyse des processus de médiatisation requiert « *de travailler à partir d'un corpus plurimédiatique portant sur l'objet étudié, corpus qui comprend l'ensemble des productions télévisuelles, radiophoniques et issues de la presse écrite, que cette dernière soit papier ou en ligne. Or, la constitution d'un tel corpus peut être compliquée à réaliser. Il s'avère en effet fastidieux d'en garantir l'exhaustivité* ». Elle souligne ici les deux points importants que sont la définition de l'objet étudié et la complétude du corpus sur les différents médias. La démarche qui est traditionnellement mise en œuvre consiste effectivement à définir l'événement que l'on souhaite étudier puis à assembler le corpus correspondant. Des approches qualitatives, quantitatives ou mixtes sont ensuite mises en

œuvre pour analyser la médiatisation de l'événement. Toutefois, Dalibert, souligne également que « *constituer ce type de corpus s'avère extrêmement intéressant lorsque les questionnements de recherche portent sur le fonctionnement de la sphère publique globale* ». L'événement va dans ce cas servir de support à une analyse qui sera plus orientée vers le système médiatique et ses modes de fonctionnement. Quel que soit l'angle choisi, l'utilisation d'un corpus plurimédiatique pose la question de l'homogénéité des documents qu'il contient. Dans notre cas, il est évident qu'un tweet est très éloigné d'une transcription de matinale radio ou d'une dépêche AFP. Pincemin (2012) aborde cette question et indique que « *l'hétérogénéité du corpus n'est pas en soi mauvaise. Elle peut être liée au terrain d'observation linguistique choisi. Il convient bien alors d'adapter les méthodes et outils aux données – et non l'inverse. Ainsi, la textométrie est, dès ses origines, attentive aux questions de normalisation et s'efforce de rester au plus près des données originales* ». Nous nous inscrivons pleinement dans cette perspective méthodologique. Nous verrons en effet que notre approche vise à identifier, via la textométrie, les documents d'intérêt au sein du corpus afin de pouvoir quantifier la proportion de la production totale qui est dédiée à la pandémie. Ce procédé permet une homogénéité de traitement sur nos sous-corpus qui sont de natures différentes.

Plusieurs études ont été réalisées sur la médiatisation du Covid et son ampleur (Petit, 2020). Elles sont issues de chercheurs, de journalistes se questionnant sur les pratiques de leur profession ou encore de médecins. La compréhension de la circulation des informations en cas d'épidémie fait en effet également partie des modélisations de propagation de la maladie (Sun, Yang et al., 2011 ; Wu, Fu et al., 2012). Mais le traitement médiatique peut également avoir des effets sur la santé de la population. C'est une des conclusions d'un article de synthèse (Prescrire, 2021) qui explique qu'« *une part importante du grand public a considéré que le traitement médiatique de la crise sanitaire a été anxiogène. Plusieurs pratiques de mise en forme de l'information, par exemple la présentation de statistiques dépourvues de contexte ou le recours à des témoignages forts de patients, ont sans doute contribué à ce ressenti. Des enquêtes dans divers pays ont suggéré un lien entre la durée d'exposition aux médias et le niveau d'anxiété ressenti, de symptômes dépressifs et de troubles du sommeil. La nature de certaines informations a aussi pu nuire à la santé mentale d'une partie de la population.* » Les travaux du géographe Claude Grasland sont ceux qui se rapprochent le plus de notre méthodologie. Il a étudié la médiatisation du Covid-19 dans la presse internationale à travers 125 journaux localisés dans 25 pays différents (Grasland, 2020 ; Grasland et Vincent, 2020). Pour la France, 24 journaux de PQR sont inclus dans l'étude. Il note qu'« *une telle synchronisation de l'agenda des grands journaux de presse internationale n'a probablement jamais été observée au cours des dernières décennies, même au moment des attentats du 11 septembre 2001. Pour la première fois également, le champ sémantique même des médias s'est trouvé complètement envahi par celui de la crise sanitaire.* » Faisant le lien, dès la fin mars 2020, entre l'ampleur de la médiatisation et son caractère anxiogène, Pierre-Carl Langlais observe également un corpus de presse et y étudie l'utilisation du terme de « *psychose* » (Langlais, 2020). Les réseaux sociaux sont évidemment le lieu d'échanges intenses sur la pandémie. Le lien entre ces discussions en ligne et la médiatisation est par exemple étudié par des chercheurs de l'EPFL et des journalistes du Temps (Rappaz, Quelled et al., 2020). Twitter (Moysan, 2020) et Facebook (Jourdain, 2020) sont particulièrement scrutés. (Moliner, 2020) a observé la circulation des tweets selon qu'ils relaient des informations issues des médias ou non, avec un focus sur Didier Raoult. Le corpus utilisé est toutefois extrêmement faible (16 000 tweets captés sur une journée). Pour l'audiovisuel, les chiffres de l'Ina dessinent la même tendance que pour les autres supports (Poels et Lefort, 2020 ; Inastat, 2020). Sur six mois, plus de la moitié des sujets des JT sont consacrés au Covid-19 et on atteint un pic de 80 % du temps d'antenne pendant le confinement sur les chaînes historiques. Outre l'amplitude et la durée de la médiatisation, les premiers mois de la pandémie ont également vu apparaître un phénomène médiatico-politique qui aura des

répercussions au niveau mondial : Didier Raoult et la chloroquine (Demagny, 2020 ; Longhi, 2020).

Corpus exhaustif de données

La particularité de notre approche est que nous captions en permanence les contenus médiatiques et ceux circulant sur Twitter en France. Cela nous offre un avantage précieux pour des études de ce type. En effet, nous évitons ainsi deux des principaux écueils de la constitution de corpus d'actualité. D'une part, nous disposons de l'antériorité des documents, il n'est donc pas nécessaire de chercher à reconstituer l'historique d'un événement médiatique au moment où la décision est prise de l'étudier. D'autre part, nous essayons de capter de façon exhaustive les contenus qui sont publiés. Nous évitons ainsi les biais potentiels (souvent via la création de requêtes adéquates) liés à la constitution même des corpus que certaines analyses peuvent, parfois, ignorer.

Les corpus assemblés pour cette étude débutent en décembre 2019 et couvrent 210 jours pour se terminer fin juin 2021. Même si aucune médiatisation sur le Covid-19 n'est évidemment disponible pour le mois de décembre, nous utiliserons cette période comme jeu de contrôle pour les méthodes quantitatives que nous appliquerons pour analyser le reste du corpus et ainsi mesurer le bruit de fond médiatique occasionné par notre approche textométrique.

Les dépêches AFP d'actualité générale ainsi que les articles de presse en ligne sont captés par une plateforme mise en place depuis plusieurs années (Hervé, 2019). Nous ne conservons que les dépêches en français qui concernent un événement en particulier. Aussi, toutes les dépêches concernant des prévisions de publication ou des agendas ne sont pas prises en compte dans cette étude. Ces dépêches spécifiques sont identifiées grâce à leurs titres qui commencent systématiquement par certains mots caractéristiques¹. Une fois ce filtrage effectué, il nous reste 200 300 dépêches AFP dans le corpus, soit une moyenne de 954 dépêches par jour. La répartition temporelle est relativement équilibrée, avec une baisse significative au moment des fêtes de fin d'année 2019 et un cycle de production classique semaine/week-end².

Nous captions 200 sites de médias en ligne à partir de leurs flux RSS et de leur Sitemap. Pour cette étude, nous avons retenu deux critères pour inclure un média dans le corpus : ne pas avoir eu de problème technique dans la captation sur la période observée³ et, surtout, avoir la capacité d'identifier les journalistes auteurs⁴ des articles. Nous avons besoin de cette information pour la dernière partie de notre étude. Au final, nous avons sélectionné 38 médias⁵ pour un total de 872 046 articles. Ce sont les principaux titres de presse nationale (quotidienne et hebdomadaire) et quelques titres de PQR et de presse thématique. Le premier tour des élections municipales a donné lieu à la publication automatique, par certains titres de presse, d'articles permettant de rendre compte des résultats dans de nombreuses communes.

¹ Exemples de début de titres de dépêches AFP ignorées : « agenda », « prévisions », « à la une », « à noter pour aujourd'hui », « en attendant demain », « le monde en bref », « l'essentiel de l'actualité ».

² Nombre moyen de dépêches par jour de la semaine : lundi 979, mardi 1099, mercredi 1106, jeudi 1125, vendredi 1077, samedi 606 et dimanche 663.

³ Les principaux problèmes concernent les changements techniques sur les sites de presse qui conduisent soit à des modifications des URL de flux RSS ou des Sitemap soit à des changements de la structure HTML des pages nécessitant parfois des adaptations de notre code qui extrait le contenu des articles.

⁴ Nous utilisons les métadonnées présentes dans différentes balises HTML standardisées (schema.org et open graph) ou dans les propriétés CSS (author) spécifiques à certains titres de presse.

⁵ Médias web captés : 20 minutes, BFMTV, Challenges, Closer, Les Dernières Nouvelles d'Alsace, Europe 1, Femme actuelle, Football, Gala, Jeux Vidéo, L'Équipe, L'Est Éclair, L'Est Républicain, L'Obs, La Croix, La Dépêche du Midi, Le Dauphiné, Le Figaro, Le Journal du Dimanche, Le Monde, Le Parisien, Le Point, Le Télégramme, Les Echos, LCI, Libération, Management, Mediapart, Midi Libre, Next INpact, Nice Matin, Numerama, Paris Match, Pure People, Sciences et Avenir, Sport24, Valeurs Actuelles et Voici.

Ces articles sont générés par des robots. Nous avons choisi de neutraliser les publications du 15 mars 2020 pour les sites de presse. Notre étude portera donc au final sur 867 199 articles de presse. Comme pour l'AFP, on observe une baisse d'activité autour de Noël 2019 et un cycle semaine/week-end. En revanche, on note une légère baisse globale de l'ordre de 15 % du nombre d'articles à partir du confinement.

Pour l'audiovisuel, nous nous focalisons dans cette étude sur les canaux d'information en continu. Nous avons inclus les 5 chaînes TV (BFMTV, CNews, LCI, franceinfo: et France 24) et la radio France Info entre 6 h et minuit chaque jour. Cela représente un volume de 3 780 heures par canal. Pour traiter ces données, nous commençons par extraire le texte de tout ce qui est prononcé à l'antenne. Cette transcription des flux audio est assurée à l'aide du logiciel développé par le LIUM⁶ (Tomashenko, Vythelingum et al., 2016). Comme tous les logiciels de ce type, la transcription est basée sur un dictionnaire de termes et sur un modèle de la langue française⁷. Les résultats obtenus sont de bonne qualité mais ne peuvent pas être parfaits. Les erreurs peuvent être dues à deux raisons principales. Il s'agit d'une part des situations dans lesquelles les conditions acoustiques sont mauvaises : la transcription d'une émission en plateau télé sera de meilleure qualité qu'un micro-trottoir ou qu'une conversation téléphonique. D'autre part, il se peut également que les termes employés ne soient pas connus du logiciel : c'est principalement le cas avec certains noms propres ou avec de nouveaux termes. Ainsi le logiciel du LIUM, avec le modèle de langue dont nous disposons, connaît bien le mot « *coronavirus* » mais pas le terme « *Covid* ». Ce dernier peut donc être transcrit sous différentes formes qui sont phonétiquement proches. En voici quelques exemples trouvés dans les résultats : *co vide*, *koweït*, *code vide*, *comite*, *covic*, *lukovic*, *aucun vide*, *coville*... Puisque les créneaux transcrits sont fixes, nous ne tenons donc pas compte de la grille de programmes, des éventuelles émissions spéciales ou de journaux télé particulièrement longs. De la même manière, les coupures de publicité ou les avant-programmes sont dans le corpus. Le traitement documentaire effectué à l'Ina sur les différentes chaînes et radio n'est pas homogène et ne permet pas une granularité suffisamment fine pour segmenter et isoler les créneaux de news de manière équivalente pour tous les médias. Nous avons donc préféré conserver une approche uniforme pour ne pas biaiser nos résultats.

Les tweets sont captés selon l'approche proposée par (Mazoyer, Hudelot et al., 2018). Le principe général est de permettre une captation continue de tweets en français en se basant sur des requêtes de mots neutres les plus couramment utilisés (*stop words*). L'avantage de cette approche est qu'elle permet de capter, via l'API⁸ fournie par Twitter, un volume de tweets suffisants pour réaliser des études statistiques sans avoir à définir au préalable des termes de recherche et en garantissant une distribution des tweets captés équivalente à celle des tweets émis. Nos estimations, selon différentes approches, nous conduisent à penser que nous captions environ 60 % des tweets en français émis sur la plateforme de microblogging. Sur la période nous avons capté 1 195 450 131 tweets, soit environ 5,6 millions par jour. Il s'agit bien de tweets en français et non uniquement de tweets émis depuis la France. Nous n'utilisons pas les informations de géolocalisation des comptes Twitter ou des tweets, trop parcimonieuses et imprécises. Il faut toutefois noter que cette estimation est valable globalement pour une longue période. Ponctuellement, notamment en cas de période de forte publication, le pourcentage de tweets captés peut être plus faible. L'API de Twitter plafonne en effet le volume de tweets qui peuvent être récupérés à un instant donné. C'est pourquoi les volumes de tweets sont donnés

⁶ Laboratoire d'Informatique de l'Université du Mans.

⁷ Un modèle de langue est un modèle statistique qui représente la distribution des séquences de mots. Il permet ainsi de déterminer la probabilité qu'un mot apparaisse après un ou plusieurs autres. Combiné avec un modèle acoustique il permet au logiciel de transcription de passer du son au texte.

⁸ Application Programming Interface : solution technique permettant à deux applications de s'échanger des données. Ici Twitter permet de récupérer ses données et offre une solution logicielle pour le faire.

dans cette étude à titre indicatif. On constate par exemple une légère baisse d'activité à Noël et un net regain après le 1^{er} de l'an. Un pic est également observé pendant le confinement. Pour ces deux pics, nous ne savons pas s'ils ont été écrêtés par Twitter ou s'ils correspondent toujours à environ 60 % des tweets émis en français. On peut toutefois raisonnablement penser qu'il y a un regain d'activité sur Twitter à ces deux périodes. Il conviendra donc de se concentrer plutôt sur les ratios de tweets qui sont plus pertinents et non biaisés.

Algorithmes de quantification automatique

Nous mettons en place différents algorithmes de quantification automatique se basant sur nos jeux de données pour extraire les informations pertinentes permettant une analyse de la médiatisation du coronavirus sur les différents supports en France. Pour cela, nous avons déterminé un ensemble de mots caractéristiques permettant de cerner certains aspects de l'épidémie et de sa médiatisation. La découverte de ces mots est réalisée à partir d'une étude de leur fréquence et de leur pertinence (utilisation de TF-IDF) et ils ont ensuite été validés manuellement. Ces mots sont répartis, de façon subjective, en groupes thématiques. La détermination des mots ainsi que leur répartition a été effectuée mi-mars 2020 et n'a pas été modifiée depuis. Elle est à nos yeux valide pour la période étudiée. Nous comptons ensuite toutes les occurrences d'apparition de ces mots dans les textes (dépêches d'agence, presse en ligne et tweets) ainsi que dans les transcriptions (flux TV/radio). Tous les textes traités sont préalablement normalisés (minuscules, suppression des accents et des caractères non alphanumériques). Cette approche textométrique simple permet déjà d'observer quelques phénomènes. Nous présentons dans le tableau 1 l'ensemble des mots utilisés. Il s'agit plus précisément des préfixes de mots. Ainsi, le préfixe *confine* permet de compter tous les mots qui commencent par *confine*, par exemple *confiner*, *confinés*, *confinement*... Deux groupes de vocabulaire sont liés aux zones géographiques chinoise et italienne qui ont été importantes dans le démarrage de l'épidémie. Les quatre autres groupes concernent le virus proprement dit, les questions médicales, l'éducation ainsi que les mesures prises pour lutter contre l'épidémie. Nous définissons également le supra-groupe *coronavirus* qui englobe *virus*, *médecine* et *mesures*. Certains aspects de l'épidémie ne sont pas couverts par notre vocabulaire, notamment toutes les questions économiques (conséquences, mesures spécifiques, continuité d'activité...) ainsi que les élections municipales.

Groupe	Liste des préfixes
chine	chine, chinois, hubei, pekin, shanghai, wuhan
education	baccalaureat, cned, college, colleges, collegien, ecole, ecoli, educ, eduqu, eleves, enseign, etudi, lycee, maternel, periscolair, professeur, scolaire, scholaris, universit
italie	italie, italien, lombardie, milan, peninsule, rome, venetie, venise
medecine	asymptomat, chercheur, chirurgi, clinique, depistag, diagnost, docteur, ehpad, hopita, hospital, immunise, immunitaire, infirmier, laboratoire, lit, malad, medecin, medica, oms, pathologie, patient, pharmac, pulmonaire, reanim, recherche, respirat, samu, sanitaire, sante, scientifique, soign, symptom, traitement, urgenc, vaccin
mesures	annul, barriere, chez vous, confine, coude, ferm, fermeture, gel, geste, hydroalcoolique, les mains, masque, quarantaine, rapatrie, report, reporte, savon, stade trois, teletravail, usage unique
virus	cas confirme, contagi, contamination, contamane, corona, covid, desinfect, dix neuf, epidemi, fievr, forme severe, formes severes, gripp, infect, pandemi, pneumonie, propag, respiratoire, sars, sras, transm, viral, virolog, virus
coronavirus	ce groupe spécial cumule tous les mots de virus, mesures et medecine

Table 1 : Termes utilisés pour la textométrie

Nous avons également mis en place un algorithme permettant, sur la base d'un vocabulaire, d'estimer le temps d'antenne dédié au traitement d'un sujet. Il a été mis en œuvre dans le cadre de précédentes études comparant la médiatisation de plusieurs événements concomitants (Poirot et Hervé, 2019 ; Labracherie et Hervé, 2019). Nous ne présentons pas ici le fonctionnement de cet algorithme, les détails sont disponibles en ligne⁹. Il s'agit globalement d'une estimation de densité d'apparition des mots du vocabulaire sur la transcription du flux audio. Intuitivement, plus on observe d'apparitions de mots sur une courte période, plus la probabilité est élevée que cette période parle de la thématique en question. Un groupe virtuel *autre* est utilisé pour modéliser les autres thématiques de l'actualité qui ne seraient pas représentées dans notre vocabulaire. Les paramètres de cet algorithme ont été déterminés et validés à la suite d'une comparaison avec un décompte humain des temps d'antenne lors de la première étude. Ce sont les mêmes paramètres qui sont utilisés depuis. Seul le vocabulaire varie. Ces résultats automatiques comportent quelques biais connus et maîtrisés. La principale différence entre les deux précédentes études et celle-ci est la durée de l'événement observé. En effet, plus un événement dure dans le temps, plus son traitement médiatique va s'attarder sur ses différents aspects et donner lieu à des choix éditoriaux et des angles de traitement de l'actualité, le téléspectateur étant censé connaître le contexte de l'intervention. Le choix du vocabulaire autour du coronavirus a donc été fait en tenant compte de cette contrainte. Nous avons cherché à distinguer les mots de vocabulaire liés spécifiquement à ce virus, au domaine médical, aux mesures prises par les autorités et enfin aux conséquences. Tous ces mots peuvent évidemment apparaître dans l'actualité en dehors du contexte spécifique de la pandémie de coronavirus. C'est la raison pour laquelle nous conservons dans nos jeux de données le mois de décembre 2019. Il sert d'étalon pour mesurer les biais et avoir une bonne idée de la marge d'erreur de nos mesures pour la suite.

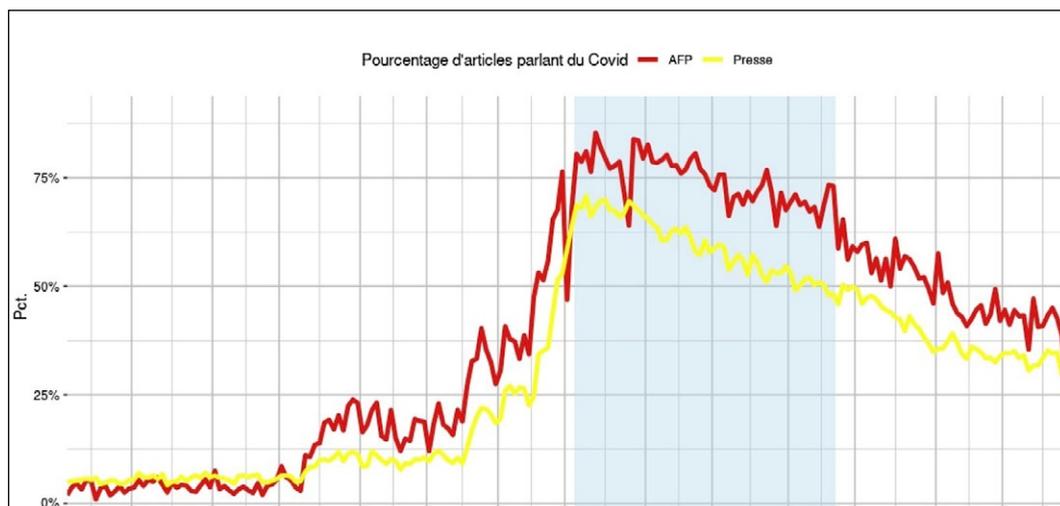


Figure 1 : Proportion d'articles de l'AFP et de la presse parlant du Covid-19

Ampleur de la médiatisation sur tous les supports

Les dépêches AFP et les articles de presse ont un titre et un contenu. Pour déterminer si un article aborde le Covid-19 nous considérons qu'il faut que cinq mots de notre vocabulaire apparaissent dans son contenu dont au moins un qui soit issu du groupe *virus*. Cette approche

⁹ Voir : www.herve.name/coronavirus

est plus stricte que celle utilisée par Grasland et Vincent (2020) qui prend en compte tous les documents dans lesquels les mots *covid* ou *coronavirus* apparaissent. On observe que quelques dépêches sont considérées comme traitant du coronavirus sur notre période de contrôle de décembre 2019. Ces faux positifs sont une bonne indication des sujets d'actualités notamment liés à la question de la gestion des hôpitaux en France avec leur manque de moyen chronique, une bactérie qui tue les orangers en Floride, la résurgence de la rougeole dans les îles Samoa ou encore d'Ebola en RDC. Ce sont ainsi très majoritairement les mots du groupe *médecine* qui sont détectés sur ce mois de décembre.

Sur les graphiques présentant les résultats, la période spécifique du confinement (17 mars au 11 mai) est signalée en bleu clair. Sur la figure 1, on observe clairement trois principales étapes dans le traitement de l'épidémie par l'AFP, avec une nette augmentation de la part de dépêches qui y est consacrée à chaque fois. La dernière semaine de janvier 2020, c'est la Chine qui est principalement traitée. On a en moyenne 17,5 % des dépêches évoquant le virus. Fin février, l'épidémie se développe fortement en Italie et on passe à un ratio de l'ordre de 31 % de dépêches liées au coronavirus. Enfin, à partir de la deuxième semaine de mars, la France est plus durement touchée et on dépasse 50 % de dépêches sur le sujet pour presque atteindre une moyenne de 74 % pendant le confinement. On remarque de plus la journée particulière du 15 mars, premier tour des élections municipales, avec un net recul de la thématique coronavirus pour ce seul jour dans les dépêches. Les périodes sont les mêmes pour les articles de presse, mais dans une proportion légèrement moindre. La diversité des natures de titres de presse que nous avons inclus dans notre corpus explique cette différence. Si on ne considère que les neuf titres qui sont assimilés à la PQN et les 251 081 articles correspondants, la courbe pour la presse se superpose alors parfaitement avec celle de l'AFP.

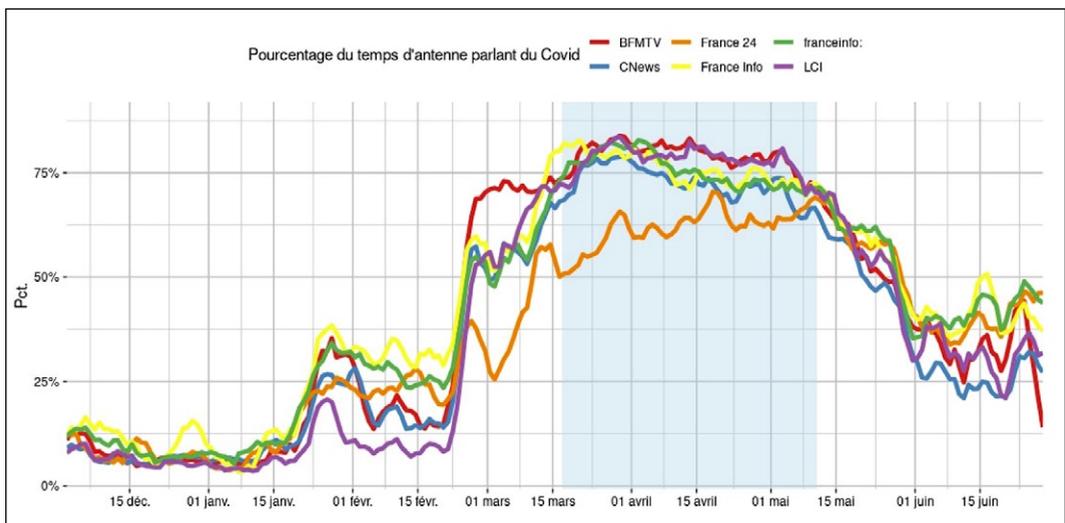


Figure 2 : Temps d'antenne consacré au Covid-19 sur les canaux d'information en continu

La figure 2 présente le pourcentage de temps d'antenne parlant du coronavirus pour les canaux d'information en continu. Pour une meilleure lisibilité, ces proportions sont calculées avec une moyenne glissante de 5 jours afin de lisser les quelques irrégularités. Contrairement à d'autres gros événements de *breaking news* tels que les attentats, les chaînes privées BFMTV, LCI et CNews n'ont pas supprimé la publicité de leurs grilles respectives. Celle-ci n'a donc pas été exclue du périmètre de l'étude, cela peut représenter une légère différence avec les chaînes publiques.

On constate que les chaînes d'information en continu ont globalement toutes un profil similaire sur cette période, proche de ceux de la presse écrite et de l'AFP. On note toutefois quelques différences. Ainsi, à la suite du premier pic de médiatisation fin janvier, la radio France Info et la télé franceinfo restent à un niveau plus élevé que les autres chaînes, supérieur à 25 % du temps d'antenne, jusqu'à l'apparition du pic suivant lors de la 3^e semaine de février. Sur ce mois, elles consacrent ainsi deux fois plus de temps d'antenne au coronavirus que LCI. Le second pic de médiatisation à partir de cette 3^e semaine de février est plus marqué en revanche sur BFMTV qui n'est rattrapé par les autres chaînes qu'à partir de mi-mars. BFMTV, centrée sur « *la priorité au direct* », et dont la ligne éditoriale et la grille présentent la plus forte élasticité à l'actualité, a embrayé beaucoup plus rapidement que ses concurrentes sur la couverture du Covid-19. Dès le lundi 24 février et pendant plusieurs semaines, la chaîne a couvert bien davantage l'actualité de ce coronavirus, suivant notamment en cela l'inflation du nombre de cas en France, et de plusieurs autres événements que nous détaillons par la suite. Seule France 24 semble se détacher et consacrer un temps d'antenne moindre au coronavirus. Le confinement marque un épisode de saturation de l'antenne qui est quasi exclusivement dédiée à la pandémie. Tout se passe en fait sur les chaînes d'information en continu comme si elles étaient en édition spéciale permanente. On remarque que le premier tour des élections municipales n'est absolument pas visible sur ce graphique, ce qui est parfaitement cohérent avec le sentiment général de celles et ceux qui étaient devant leur poste ce soir-là. Si l'on regarde maintenant le vocabulaire, on note une légère baisse des termes directement liés au *virus* depuis début mars. Comme pour Twitter, le contexte étant maintenant bien installé, il n'est plus la peine de répéter aussi souvent à l'antenne les mots du coronavirus. Les téléspectateurs ont bien intégré cela. En revanche, on observe sur la même période une montée des termes *medecine* et *mesures*.

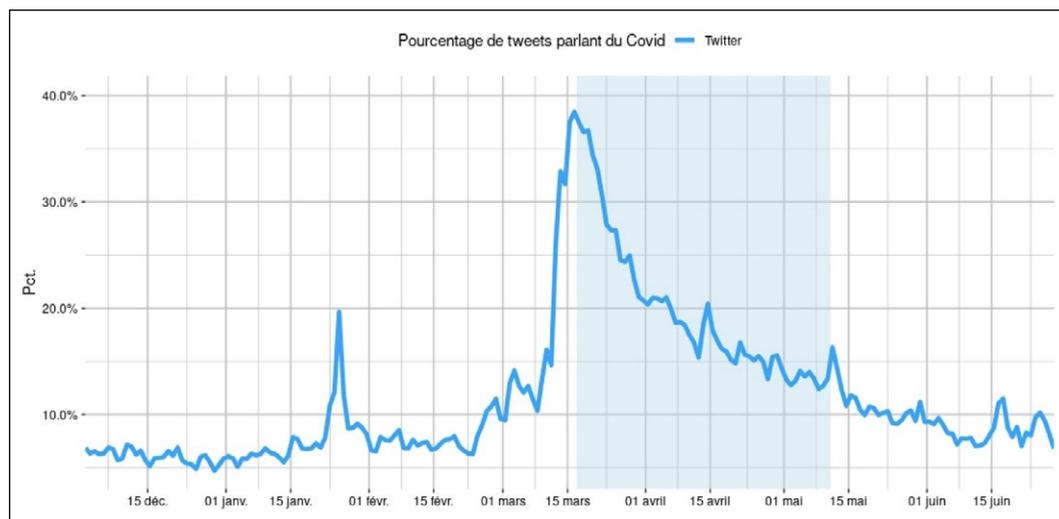


Figure 3 : Proportion de tweets consacrés au Covid-19

Par rapport aux dépêches AFP, on observe sur Twitter un comportement beaucoup plus marqué. Sur la figure 3, un premier pic apparaît fin janvier avec la médiatisation de la situation en Chine et le premier cas français puis retombe rapidement. Un léger palier est franchi début mars, mais c'est ensuite clairement à partir du 12 mars et de la première allocution télévisée d'Emmanuel Macron sur le sujet qu'un cap est franchi avec environ 35 % des tweets en français qui évoquent le sujet. La décroissance pendant le confinement est beaucoup plus marquée que sur les médias d'information. Si on regarde plus précisément les groupes de vocabulaire, on

remarque depuis le 15 mars une baisse de l'utilisation des termes liés au virus et une nette augmentation de ceux liés aux mesures prises et, dans une moindre mesure, des termes médicaux. Les rumeurs de confinement circulent en effet pendant ce week-end d'élections et ce dernier est effectivement annoncé le lundi 16. Le contexte étant maintenant évident pour tout le monde, il n'est peut-être plus nécessaire d'évoquer dans les tweets (où la place manque parfois) les termes liés au virus, et les récits du confinement deviennent plus présents. La situation dans les hôpitaux est également un sujet de discussion. La fermeture des établissements scolaires provoque un léger pic sur ce vocabulaire mais il retombe rapidement.

Chronologie et principaux déclencheurs de la médiatisation de la pandémie

Nous regroupons sur la figure 4 les résultats quantitatifs des quatre types de média que nous avons analysés et y ajoutons les principaux faits marquants liés à la pandémie sur la même période. Pour les télévisions et radios d'information en continu, nous avons simplement effectué la moyenne des six canaux étudiés. Cette chronologie nous permet de distinguer quels sont les moments clés lors desquels les courbes d'intensité médiatique marquent une inflexion nette, traduisant ainsi des choix éditoriaux de la part des rédactions dans le traitement de l'actualité ou des discussions sur Twitter.

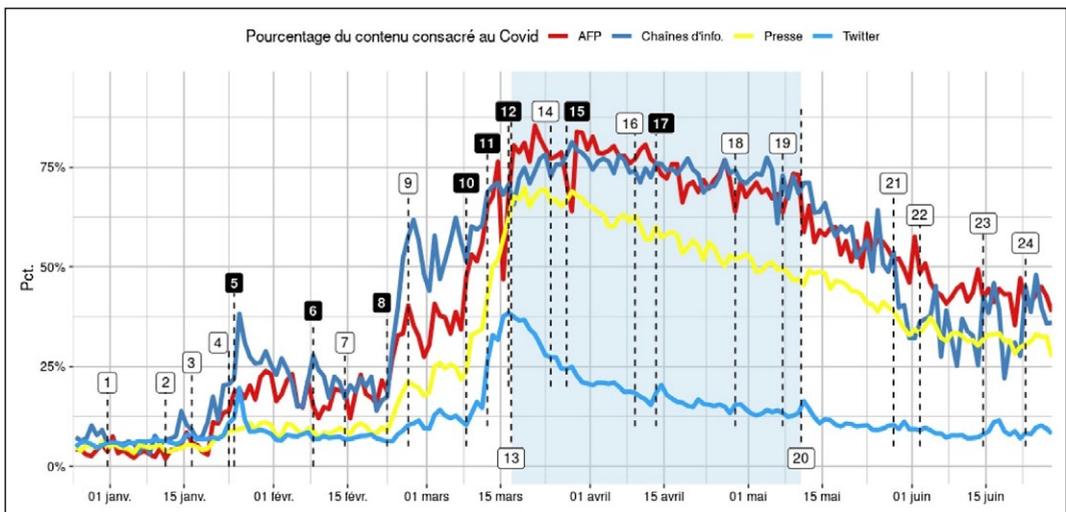


Figure 4 : Proportion de tweets consacrés au Covid-19

Si l'on reprend la *timeline* de propagation de la pandémie de Covid-19 provoquée par le coronavirus SRAS-CoV-2, le premier pic est assez clair : il concerne l'annonce des premiers cas de coronavirus testés et traités en France. Vendredi 23 janvier, à 20 h 16, Agnès Buzyn, alors ministre de la Santé, avait ainsi confirmé l'identification de deux cas en France, l'un à Bordeaux, l'autre à Paris. L'information avait été largement traitée le samedi 24 janvier dans les médias audiovisuels, plusieurs conférences de presse, notamment de soignants de l'hôpital Bichat, où l'un des cas était traité, ayant même été diffusées en direct ce jour. Un léger rebond est ensuite observé à la fin de la première semaine de février avec la découverte de plusieurs Anglais contaminés à Contamines-Montjoie et ce qui va devenir le premier foyer épidémique de France. Après une période de stagnation, le début du deuxième pic s'établit le 22 février. Plusieurs événements expliquent cette véritable explosion dans le temps d'antenne consacré au sujet. Il s'agit, d'une part, de l'annonce du premier mort français, mercredi 26 février : un homme de 60 ans, originaire de l'Oise, et dont l'origine de la contamination est alors présentée

comme un mystère. Un premier mort en France avait été recensé le 14 février, un homme de 80 ans originaire de Chine, mais l'information n'avait pas provoqué d'augmentation détectable dans nos analyses. L'autre information qui explique le renforcement du temps d'antenne à partir du 24 février concerne l'Italie, et la mise en place de plusieurs mesures de protections sanitaires voire d'isolement. Nous avons ainsi déterminé un vrai pic dans l'emploi de termes liés à l'Italie au début de la semaine du 24 février.

31 déc.	OMS : pneumonie atypique en Chine	17 mar.	Début du confinement
11 jan.	Premier mort de l'épidémie en Chine	24 mar.	Seuil des 1 000 décès à l'hôpital franchi
16 jan.	Arrivée en France du premier cas avéré	27 mar.	Annonce prolongation du confinement
23 jan.	Wuhan confinée	09 avr.	E. Macron à l'IHU de Marseille
24 jan.	Premiers cas à Paris et Bordeaux	13 avr.	E. Macron 20 h, prolongation du confinement
08 fév.	5 Anglais contaminés en Haute Savoie	28 avr.	E. Philippe dévoile plan déconfinement
14 fév.	Le premier décès en France	07 mai	E. Philippe annonce le déconfinement
22 fév.	L'Italie isole une dizaine de villes	11 mai	Fin du confinement
26 fév.	Premier mort français dans l'Oise	28 mai	Annonce 2 ^{de} phase du déconfinement
08 mar.	Cap des 1 000 cas franchi en France	02 juin	Réouverture bars, restaurants, écoles
12 mar.	E. Macron 20 h, fermeture des écoles	14 juin	Macron s'exprime en direct
16 mar.	E. Macron 20 h, annonce du confinement	22 juin	Début de la phase 3 du déconfinement

Table 2 : Éléments de la chronologie médiatique du Covid-19

Le bilan éditorial de cette séquence sur le temps long est clair : il s'agit d'une illustration très nette de la loi du « *mort-kilomètre* ». Cette règle voudrait que le public s'intéresse davantage à une personne qui meurt au coin de sa rue qu'à la mort de 3 000 personnes à l'autre bout du monde. Cette lecture de l'actualité par le prisme de la proximité de l'audience au sujet s'est exprimée sur la période au moins à deux reprises : lors du premier pic de médiatisation, avec l'annonce des deux premiers cas en France, le samedi 25 janvier ; mais il se retrouve même avec une variante dans les semaines qui suivent : la mort d'un touriste chinois à Paris le 14 février. Quand bien même il s'agit d'un événement très important sur le plan médical, puisqu'il s'agit du premier mort hors du continent asiatique, cet événement génère une couverture médiatique beaucoup moins intensive que la mort d'un Français, le mercredi 26 février. Autrement dit, la loi du « *mort-kilomètre* » peut encore être précisée : une personne étrangère qui meurt à proximité intéressera moins que la mort d'un Français intervenant au même endroit. Ensuite, les faits s'enchaînent plus rapidement pour aboutir à la saturation de l'espace médiatique. Le franchissement du cap symbolique des 1 000 cas en France le 8 mars est le dernier coup d'accélérateur avant les deux interventions télévisées très rapprochées du Président de la République des 12 et 16 mars. On entre alors dans la phase du confinement. Pendant cette période, la presse et les médias audiovisuels étant déjà quasi exclusivement focalisés sur l'épidémie, les deux annonces de prolongation du confinement ne sont pas perceptibles. Elles provoquent toutefois un rebond visible dans les conversations du Twitter. Enfin, en fin de période analysée, on observe un décrochement des médias audiovisuels à partir du 28 mai et de l'annonce de la deuxième étape du déconfinement.

Didier Raoult et la Chloroquine : le rôle des politiques dans la médiatisation

Nous avons également mesuré, sur les différents supports, l'apparition et l'importance consacrée au professeur Didier Raoult, qui a assuré, dans deux études controversées publiées les 20 et 27 mars, avoir trouvé un traitement efficace contre le Covid-19, la chloroquine. L'apparition dans le débat de ce professeur de médecine, directeur de l'Institut Hospitalo-

Universitaire Méditerranée Infection (IHU) à Marseille, constitue clairement un feuilleton médiatico-politique sur les premiers mois de la pandémie qui a eu des répercussions au niveau mondial, parfois dramatiques.

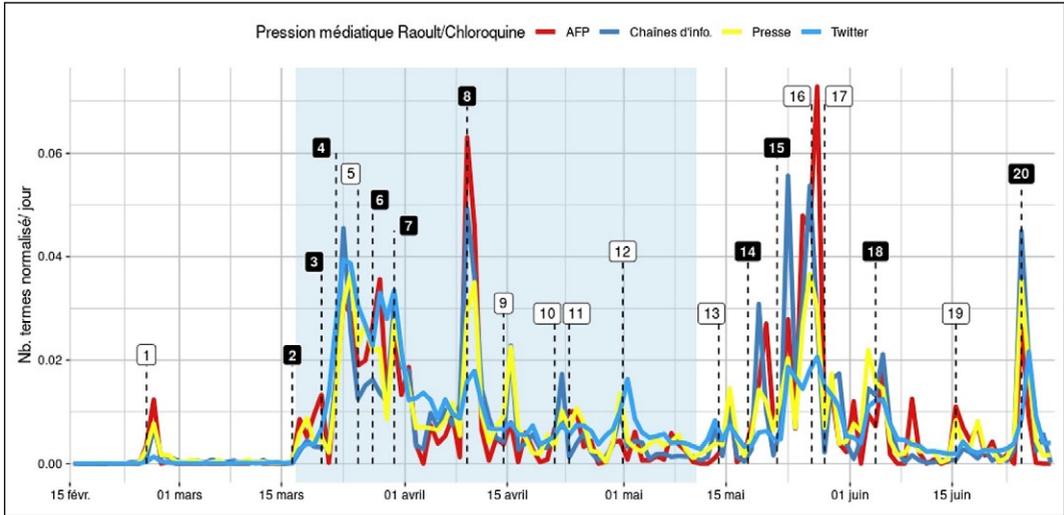


Figure 5 : Pression médiatique sur les différents supports du phénomène Raoult/Chloroquine

Nous utilisons un vocabulaire simplifié¹⁰ pour regarder la médiatisation liée à la chloroquine. Contrairement aux autres quantifications effectuées sur les données Twitter, nous avons ici conservé les mentions afin de ne pas invisibiliser le compte Twitter *@raoult_didier*. Nous n'utilisons que l'approche textométrique en comptant le nombre d'occurrences des termes, sans chercher à quantifier le temps d'antenne car isoler ce sous-événement des autres est trop délicat puisqu'il est évidemment traité dans le contexte global de la pandémie. Les courbes de la figure 5 présentent ce nombre d'occurrences, normalisée par support. On ne peut pas comparer les ampleurs relatives des supports mais on peut observer les moments d'inflexion de ces courbes.

25 févr.	D.R. « <i>Fin de partie pour le coronavirus</i> »	23 avr.	Agence europ. médicaments met en garde
16 mar.	D.R. « <i>Résultats positifs de ma 1^{re} étude</i> »	30 avr.	Didier Raoult BFM TV/Paris Match
20 mar.	Donald Trump a « <i>approuvé la chloroquine</i> »	14 mai	Inefficacité HCQ selon plusieurs études
22 mar.	Retailleau, Estrosi soutiennent HCQ	18 mai	Trump prend de l'HCQ
24 mar.	D.R. quitte le conseil scientifique	22 mai	Étude du <i>Lancet</i>
27 mar.	L'appel de Philippe Douste Blazy	26 mai	D.R. sur LCI
30 mar.	Le Pen veut que les médecins puissent prescrire	28 mai	100 médecins : dix travers méthodologiques
09 avr.	Macron à l'IHU, 3 ^e étude	04 juil.	Retrait article du <i>Lancet</i>
14 avr.	L'AP-HP lance une étude clinique	15 juil.	FDA révoque autorisation en urgence HCQ
21 avr.	Le NIH déconseille l'HCQ	24 juil.	Raoult auditionné commission enquête AN

Table 3 : Éléments de la chronologie médiatique Raoult/Chloroquine

¹⁰ Termes utilisés : raoult, didier raoult, ihu, chloroquine, nivaquine, hydroxychloroquine, plaquenil.

Le premier gros pic se produit entre mi-mars et début avril, au début du confinement. Il est amorcé par l'annonce de Didier Raoult de sa première étude et de ses résultats supposément positifs. Mais l'emballement a lieu quelques jours plus tard avec les déclarations de Donald Trump et son souhait de mettre en avant le traitement à base de chloroquine. La brèche est ouverte et la droite et l'extrême droite françaises se saisissent de l'occasion pour mettre la pression sur le gouvernement en soutenant très ouvertement Didier Raoult. Ces 15 jours sont également le pic de discussions sur Twitter, avec un niveau qui ne sera plus atteint pas la suite. La première semaine d'avril est relativement calme sur le sujet puis la sortie de la troisième étude de Didier Raoult, combinée à la visite surprise d'Emmanuel Macron à l'IHU à Marseille vont finir de crédibiliser les travaux sur la chloroquine avec un très net pic médiatique. La communauté scientifique, avec sa temporalité qui n'est pas celle des médias, commence à produire des résultats sur la deuxième moitié du confinement. Mais les premières alertes et mises en garde contre l'utilisation de la chloroquine, bien qu'émanant d'autorités publiques reconnues, sont loin de provoquer le même niveau de couverture médiatique que les études du professeur Raoult. La médiatisation reprend le 18 mai, de nouveau avec l'intervention d'un homme politique mais cette fois-ci outre-Atlantique : Donald Trump annonce qu'il prend de l'hydroxychloroquine. Quelques jours plus tard, la célèbre revue médicale *The Lancet* publie un article soulignant les dangers du traitement à base de chloroquine. Mais rapidement des critiques sont émises contre cette étude, y compris de la part de la communauté scientifique. Cette période de dix jours est le plus gros pic médiatique sur la période analysée. Le retrait de la publication de *Lancet* le 4 juin n'aura qu'un impact médiatique limité. Point d'orgue de la médiatisation de Didier Raoult, son audition en commission d'enquête parlementaire à l'Assemblée nationale le 24 juin.

Sujet de controverse très disputé sur les réseaux sociaux (Bayet et Hervé, 2020 ; Smyrnaioi, Tsimboukis et al., 2021), c'est bien son lien ambivalent avec le monde politique qui assied la notoriété de Didier Raoult. Comme le souligne (Rahmil, 2020), le premier acte qui a fait connaître Didier Raoult est un « *claquage de porte* ». L'infectiologue a quitté le conseil scientifique le 24 mars 2020, un jour après avoir créé son compte Twitter. Persuadé que son cocktail médicamenteux d'hydroxychloroquine et d'azithromycine permettrait d'éviter un confinement, il va marquer son positionnement à l'aide d'un tweet : « *Je suis en contact avec le Ministère et avec le Président de la République pour leur dire ce que je pense. Je reste en contact avec eux directement, car le conseil ne correspond pas à ce que je pense devoir être un conseil stratégique.* » Les médias se sont par la suite interrogés sur l'emballement qu'ils ont contribué à créer. Mais certains de leurs travers (tendance trop marquée à la personnalisation, faiblesse de la culture scientifique dans les rédactions, recherche de l'audience) n'ont pas pu résister au fort soutien d'une partie de la classe politique. Les principaux pics médiatiques observés sont tous directement corrélés à des interventions politiques, en France et aux États-Unis. Nous ne pouvons pas en tirer de conclusion de causalité¹¹, mais le comportement de la classe politique peut sérieusement être questionné sur ce sujet.

Impact de la pandémie sur le travail des journalistes

Le travail des journalistes s'est trouvé fortement perturbé pendant le premier confinement. D'une part, les rédactions ont dû s'adapter, comme toutes les autres activités sociales et économiques, pour intégrer les contraintes logistiques imposées pour limiter la propagation du virus. D'autre part, des pans entiers du pays étant à l'arrêt, ce sont les sujets même sur lesquels les journalistes travaillent habituellement qui se retrouvent sans actualité particulière à couvrir. On peut en avoir un premier aperçu en observant les thématiques des dépêches AFP. La production de l'agence, en volume, n'a pas particulièrement été impactée par le confinement.

¹¹ Il nous faudrait pour cela d'autres événements similaires pour pouvoir en faire une modélisation correcte, ce qui n'est clairement pas souhaitable.

Chaque dépêche est catégorisée thématiquement en suivant la classification standardisée de l'IPTC¹². Nous représentons les proportions respectives des 17 catégories principales sur le graphique 6. Le total peut dépasser 100 % puisque plusieurs catégories peuvent être attribuées à chaque dépêche.

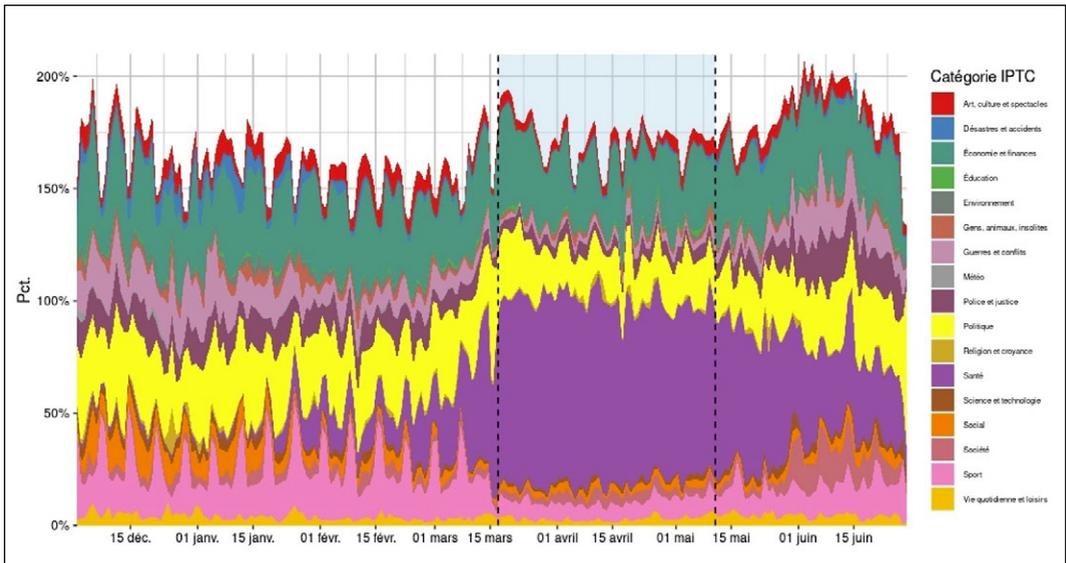


Figure 6 : Catégories IPTC des dépêches AFP

On observe naturellement la catégorie *Santé*, quasi inexistante avant la mi-janvier 2019, prendre une place prépondérante dans les dépêches. La tendance est la même que pour les sujets de JT observée par Poels et Lefort (2020). *Économie et finances* ainsi que *Politique* sont globalement stables. En revanche, on observe pour cinq catégories (*Sport*, *Social*, *Police et justice*, *Guerres et conflits*, *Désastres et accidents*) un très net ralentissement qui correspond à la période du confinement. Pour le sport, la justice ou les accidents, il peut sembler évident que la mise au pas de la vie sociale a asséché les événements que les journalistes auraient en temps normal été amenés à couvrir. En revanche pour les guerres, on s'interroge sur la diminution effective des conflits armés pendant cette période ou sur l'impossibilité pour les journalistes d'être présents pour les couvrir. À la fin du confinement, certaines catégories se retrouvent en proportion plus importantes qu'en temps normal, comme si un rattrapage s'opérait. C'est ainsi le cas des questions de *Société* mais également *Police et justice* et *Guerres et conflits*.

Une autre façon d'observer l'impact du confinement sur le travail des journalistes est de voir dans quelle mesure leur production individuelle a été significativement modifiée. Pour cela, nous allons regarder les articles des 38 médias en ligne que nous avons captés. Nous pouvons extraire les noms des auteurs depuis les métadonnées présentes dans les pages HTML. Nous ne conservons que les auteurs correctement identifiés, sans tenir compte des articles non signés ou attribués à la rédaction ou à une agence de presse. On s'intéresse aux 55 jours du confinement (période CF). On va pour cela quantifier la production des journalistes sur cette période ainsi que sur les 55 jours qui précèdent (période P1) et les 55 qui suivent (période P2). Sur ces 165 jours, soit du 22 janvier au 4 juillet, nous avons 746 172 articles au total et 249 787 pour lesquels nous pouvons identifier de façon correcte un ou plusieurs journalistes

¹² International Press and Telecommunications Council, organisme chargé des standards d'échange de données pour la presse.

en tant qu'auteurs. La production des médias ainsi que la proportion d'articles conservés sont très variables d'un média à l'autre. Cela était prévisible puisque notre corpus contient aussi bien des médias ayant de grosses rédactions que des médias de niche. De plus, leur stratégie éditoriale, notamment sur la publication de dépêches d'agences, est très variable. On conserve ainsi presque tous les articles de *Pure People* ou *Numerama*, environ 63 % pour *Le Monde*, 26 % pour *Libération*, 13 % pour *Mediapart* et 8 % pour *Challenges*. On ne va donc pas comparer les médias mais plutôt observer quelles sont les rédactions pour lesquelles de nombreux journalistes ont eu une forte variabilité de productivité sur les 3 périodes de 55 jours retenues. Pour qu'un journaliste soit inclus dans l'étude, il doit avoir publié au moins un article en P1 et P2 et avoir au moins 20 articles sur les 165 jours. Cela permet notamment d'exclure les tribunes d'opinion pour lesquels les auteurs ne font pas partie de la rédaction. Nous avons ainsi conservé 2 991 journalistes dans notre corpus. Globalement, sur les 38 médias, la distribution des articles est de 34,7 % publiés en P1, 32 % pendant le confinement et 33,3 % en P2. En utilisant un test statistique (K_{hi}²), on peut comparer chaque média à cette distribution moyenne. Les trois médias dont la distribution de publication d'articles est la plus proche de cette moyenne sont *Pure People*, *Le Monde* et *Paris Match*. À l'inverse, les 3 qui s'en éloignent le plus sont *Closer*, *Football* et *Sport24*. Si maintenant on regarde individuellement chaque journaliste, on peut également voir dans quelle mesure sa production diverge de la moyenne. Il n'est pas question ici de rapporter des statistiques individuelles tant leur interprétation ne présente pas d'intérêt particulier et surtout recouvre potentiellement des explications diverses. Nous choisissons donc de voir comment se répartit dans les rédactions le premier tiers¹³ des journalistes ayant un profil éloigné de la moyenne. Cela nous donne une proportion de l'équipe de rédaction de chaque média qui est concernée. Les 5 médias qui ont la plus forte proportion de journalistes ayant changé leurs habitudes de production sont *Football* (71 %), *Sport24* (65 %), *L'Équipe* (59 %), *Closer* (59 %) et *Gala* (53 %). Les rédactions les plus stables sont *Les Echos* (11 %), *Management* (11 %), *La Croix* (13 %), *Le Journal du Dimanche* (14 %) et *Challenges* (16 %). Pour les médias évoqués précédemment, *Le Monde* voit 17 % de ses journalistes concernés, *Pure People* 20 % et *Paris Match* 29 %. Les rédactions les plus bouleversées par le confinement concernent donc les médias sportifs et la presse people. Il faudrait une étude plus poussée sur le statut des journalistes des rédactions, notamment pour connaître la proportion de pigistes dont on sait qu'ils ont particulièrement souffert du manque de commande pendant le confinement (Dassonville, 2020). L'écart entre *Pure People* d'un côté et *Gala* et *Closer* de l'autre pourrait ainsi être envisagé. On peut toutefois raisonnablement penser que les thèmes spécifiques couverts par ces médias et la quasi-disparition de toute activité publique sur leurs sujets sont la raison principale qui a bouleversé ces rédactions.

Conclusion

Nous avons mis en place une approche quantitative sur un corpus complet couvrant les 6 premiers mois de la pandémie de Covid-19 dans l'espace médiatique français : AFP, presse en ligne, chaînes télévisées et radio d'information en continu et Twitter. Avec une approche textométrique incluant le calcul du temps d'antenne pour les médias audiovisuels, nous avons estimé la proportion de la production de contenu de tous les supports qui est dédiée à la pandémie. Nous avons ainsi déterminé les trois principales phases de l'accroissement de cette médiatisation qui aboutissent à une quasi-saturation pendant le confinement avec un taux de couverture qui frôle les 80 % pour les médias traditionnels. N'y a-t-il plus d'actualité en dehors du coronavirus à cause du confinement et de la mise à l'arrêt de la vie sociale ou bien n'y a-t-il plus d'espace médiatique disponible pour en parler ? En observant la chronologie des faits, nous notons que les principaux déclencheurs pour l'augmentation de la médiatisation sont l'apparition des premiers cas en France, la décision de l'Italie de mettre une dizaine de villes

¹³ Proportion choisie arbitrairement, en étant plus restrictif l'ordre des 5 médias les plus touchés ne change pas.

sous cloches en Lombardie suivie de peu par le premier mort français et les interventions télévisées du Président de la République. Un premier focus sur Didier Raoult fait apparaître le rôle d'une partie de la classe politique dans l'emballage médiatique qui s'est amorcé pendant le confinement. Un second focus observe les changements dans la production des journalistes de l'AFP et de la presse en ligne. Les médias sportifs et la presse people semblent être ceux pour lesquelles les rédactions ont le plus été déstabilisées par le confinement et ses contraintes.

Ce travail quantitatif ne peut en aucun cas expliquer la complexité et l'interconnexion des différents acteurs qui entrent en jeu dans la médiatisation d'un événement tel que la pandémie de Covid-19. Tout au plus pouvons-nous en quantifier l'amplitude et mettre quelques chiffres sur des phénomènes qui sont déjà connus¹⁴. Les statistiques descriptives que nous avons produites doivent toutefois être utiles à d'autres travaux, notamment ceux qui choisissent une approche qualitative en information-communication ou en sociologie pour mieux comprendre le fonctionnement des salles de rédaction et les choix des journalistes dans des situations aussi atypiques que celles récemment vécues. Du côté quantitatif, nous sommes persuadés que le travail sur des corpus les plus complets possibles, même si cela implique de lourdes contraintes d'ingénierie, est celui qu'il faut privilégier pour limiter un certain nombre de biais. Nos futurs travaux s'orientent vers de nouveaux algorithmes permettant d'atteindre une granularité plus fine dans la description des événements médiatiques afin de pousser l'observation à des niveaux plus précis. ■

Nicolas Hervé est chercheur au Service de la recherche de l'Institut National de l'Audiovisuel.

Références

- Bayet, Antoine et Nicolas Hervé (2020). ÉTUDE INA. Comment Didier Raoult et la chloroquine ont surgi dans le traitement médiatique du coronavirus. *La revue des médias* [en ligne] larevuedesmedias.ina.fr, 31.03.2020.
- Dalibert, Marion (2018). Reconstituer l'événement. Dans Sarah Lécossais et Nelly Quemener (dirs), *En quête d'archives : bricolages méthodologiques en terrains médiatiques* (p. 69-77). INA Publications.
- Dassonville, Aude (2020). Les journalistes pigistes paient le prix fort de la crise dans la presse. *Le Monde* [en ligne] lemonde.fr, 05.06.2020.
- Demagny, Xavier (2020). YouTube, Twitter, Facebook : Didier Raoult est devenu une star du web (et pas que pour le meilleur). *France Inter* [en ligne] radiofrance.fr, 27.03.2020.
- Grasland, Claude (2020). Comment la pandémie s'est propagée dans la presse régionale. *The Conversation* [en ligne] theconversation.com, 19.08.2020.
- Grasland, Claude et Jean-Marc Vincent (2020). Much ado about one (single) thing?: H2020 ODYCCCEUS internal seminar. [En ligne] *HAL-SHS*, 05.06.2020.

¹⁴ Cette étude est la première qui combine tous les flux captés et analysés par la plateforme de recherche OTMedia. Ce projet fut imaginé et initié par la regrettée Marie-Luce Viaud, cette publication lui est dédiée. Nos remerciements vont à Thomas Dugeon, Fabien Larrieu, Antoine Laurent, Pierre Letessier, Béatrice Mazoyer, Sylvain Meignier, Richard Poirot, Denis Rakulan, Agnès Saulnier et Denis Teyssou pour avoir permis, par leurs apports respectifs, de faire en sorte que les briques technologiques et les données soient disponibles pour cette étude.

- Hervé, Nicolas (2019). OTMedia, the TransMedia News Observatory. *FIAT/IFTA Media Management Seminar 2019*.
- INA (2020). Le baromètre thématique des journaux télévisés n° 58 : la Covid-19 dans les JT. [En ligne] *inatheque.fr*, 09.2020.
- Jourdain, Stéphane (2020). Selon une étude, 94 % des commentaires Facebook portent sur le coronavirus. *France Inter* [en ligne] *radiofrance.fr*, 20.03.2020.
- Labracherie, Juliette et Nicolas Hervé (2019). ÉTUDE. Incendie de l'usine Lubrizol à Rouen et mort de Jacques Chirac : comment les chaînes d'info ont traité d'une double actualité. La revue des médias [en ligne] *larevuedesmedias.ina.fr*, 07.10.2019.
- Langlais, Pierre-Carl (2020). De la « psychose » à la crise : le coronavirus dans la presse française. *Sciences communes* [en ligne] *scoms.hypotheses.org*, 26.03.2020.
- Longhi, Julien (2020). Le vrai problème Raoult : quand les médias transforment la science en spectacle. *HuffPost* [en ligne] *huffingtonpost.fr*, 11.06.2020.
- Mazoyer, Béatrice, Céline Hudelot, Marie-Luce Viaud et Julia Cagé (2018). Real-time collection of reliable and representative tweets datasets related to news events. Dans *BroDyn 2018: Workshop on analysis of broad dynamic topics over social media*, 03.2018.
- Moliner, Pascal (2020). Médias, relais et discussions sur Twitter : proximités et distances lexicales à propos du Covid-19. *Communication & Organisation*, 58, 89-107.
- Moysan, Thomas (2020). Mathilde Guinaudeau (Ipsos) : « Twitter concentre 75 % des conversations en ligne sur le coronavirus en France ». [En ligne] *cbnews.fr*, 26.03.2020.
- Petit, Cyril (2020). Coronavirus : près de 19.000 articles chaque jour dans la presse française, un record. *Le Journal du Dimanche* [en ligne] *lejdd.fr*, 21.03.2020.
- Pincemin, Bénédicte (2012). Hétérogénéité des corpus et textométrie. *Langages*, 187, 13-26.
- Poels, Géraldine et Véronique Lefort (2020). Covid-19 dans les JT : un niveau de médiatisation inédit pour une pandémie. *La revue des médias* [en ligne] *larevuedesmedias.ina.fr*, 01.10.2020.
- Poirot, Richard et Nicolas Hervé (2019). Les « gilets jaunes », trou noir médiatique. *La revue des médias* [en ligne] *larevuedesmedias.ina.fr*, 30.07.2019.
- Prescrire, rédaction (2021). Covid-19 : l'exposition aux médias d'information, possible facteur d'anxiété. *Prescrire*, 41(455), 690-695, 01.09.2021.
- Rahmil, David-Julien (2020). Didier Raoult : itinéraire d'un animal médiatique. *L'ADN* [en ligne] *ladn.eu*, 06.11.2020.
- Rappaz, Jérémie, François Quéllec et Paul Ronga (2020). Covid-19 : histoire d'une médiatisation. *Le Temps* [en ligne] *letemps.ch*, 20.03.2020.
- Smyrnaiois, Nikos, Panos Tsimboukis et Lucie Loubère (2021). La controverse de Didier Raoult et de sa proposition thérapeutique contre la COVID-19 sur Twitter : analyse de réseaux et de discours. *Communiquer*, 32, 63-81.
- Sun, Chengjun, Wei Yang, Julien Arino et Kamran Khan (2011). Effect of media-induced social distancing on disease transmission in a two patch setting. *Mathematical Biosciences*, 230(2), 87-95.

Tomashenko, Natalia, Kévin Vythelingum, Anthony Rousseau et Yannick Estève (2016). LIUM ASR systems for the 2016 Multi-Genre Broadcast Arabic Challenge. Dans *2016 IEEE Spoken Language Technology Workshop* (p. 285-291). IEEE.

Wu, Qingchu, Xinchu Fu, Michael Small et Xin-Jian Xu (2012). The impact of awareness on epidemic spreading in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(1).