







*L'information à tout prix*

COLLECTION « MÉDIAS ET HUMANITÉS »

- De Gaulle et la diplomatie par l'image*, Olivier Chantriaux  
*L'écrivain et son public à l'ère de la radio, D'Edgar Allan Poe à Paul Valéry*, Jane Blevins  
*Le poste à transistors à la conquête de la France. La radio nomade (1954-1970)*, Elvina Fesneau  
*Du service public à la télé-réalité. Une histoire culturelle des télévisions européennes*, Jérôme Bourdon  
*Carbone 14. Légende et histoire d'une radio pas comme les autres*, Thierry Lefebvre  
*Jean d'Arcy, penseur et stratège de la télévision française*, Sylvie Pierre  
*Histoire de la télévision régionale. De la RTF à la 3, 1950-2012*, Benoit Lafon  
*«Le chantre de l'opinion». La communication de Michel Rocard de 1974 à 1981*, Pierre-Emmanuel Guigo  
*12 ans de télé-réalité... au-delà des critiques morales*, Nathalie Nadaud-Albertini  
*Frédéric Mitterrand. Une approche du 7<sup>e</sup> art au petit écran*, Juliette Courtois  
*Les Trente Glorieuses du téléspectateur. Une histoire de la réception télévisuelle des années 1950 aux années 1980*, Géraldine Poels  
*La propagande radiophonique nazie*, Muriel Favre  
*Ignorée de tous... sauf du public*, Noël Burch et Geneviève Sellier  
*Les Shadoks. Histoire, esthétique et 'pataphysique*, Sébastien Denis  
*Silences et récits, les médias français à l'épreuve des conflits africains (1994-2015)*, François Robinet  
*La cuisine en spectacle, les émissions de recettes à la télévision française (1953-2012)*, Olivier Roger  
*Les émissions de variétés de Maritie & Gilbert Carpentier (1948-1988), un divertissement français de la seconde moitié du XX<sup>e</sup> siècle*, Elizabeth Jacquinet

L'image de la couverture a été réalisée  
grâce au site [nuagesdemots.fr](http://nuagesdemots.fr)

© INA Éditions 2017  
94360 Bry-sur-Marne

ISBN : 978-2-86938-248-0

Julia Cagé, Nicolas Hervé  
et Marie-Luce Viaud

# L'information à tout prix

INA

### *Remerciements*

Les auteurs tiennent à remercier très chaleureusement Agnès Chauveau pour avoir cru dès le départ en ce projet et fait en sorte que ce livre voie le jour, et Géraldine Poels pour avoir géré sa mise en œuvre.

Karen Bastien, Brice Terdjman et Clément Thorez de WeDoData ont réalisé les graphiques de ce livre et ainsi contribué à rendre plus explicites et lisibles nos résultats et nos données. Nous leur sommes infiniment reconnaissants pour leur patience et leur formidable travail. Ce livre ne serait pas le même sans leur contribution !

Ce livre n'existerait pas sans Agnès Saulnier, Jérôme Thièvre et Thomas Drugeon qui ont participé au projet OTMedia. Nous les remercions infiniment pour leur aide tout au long de nos travaux.

Nos réflexions sur la protection de la propriété intellectuelle en France et aux États-Unis doivent beaucoup à Benoît Huet qui nous a éclairés lorsque que nous étions perdus dans les complexités des systèmes juridiques des deux côtés de l'Atlantique. Nous lui en sommes extrêmement reconnaissants.

Enfin, tous nos remerciements vont à Agathe, Sabrina et Thomas pour leurs (parfois multiples !) relectures attentives et leurs nombreuses remarques pertinentes. Ce livre est le fruit des riches échanges et discussions que nous avons eus ensemble, et doit beaucoup à leur soutien sans faille durant les longs mois d'écriture.

*Avertissement*

Ce livre présente les résultats d'une collaboration scientifique entamée dès 2012. Julia Cagé est membre du Conseil d'administration de l'AFP depuis novembre 2015. Aucun des résultats n'a été influencé d'une manière ou d'une autre par cette nomination. De plus, les opinions présentées dans ce livre sont celles des seuls auteurs. Elles ne reflètent en aucune manière le point de vue de l'AFP ou de l'INA.



## INTRODUCTION

Remède à tous leurs maux ou fossoyeurs des médias traditionnels? Formidable outil de démocratisation de l'information ou vecteur d'uniformisation? Producteur d'information originale ou lanceur de rumeurs? En l'espace de deux décennies, Internet est venu bousculer la production et la consommation d'information avec tant de force qu'en comparaison, la radio et la télévision apparaîtraient presque comme des innovations médiatiques mineures. Confrontés à une telle révolution, depuis 1993 et la naissance du web, depuis 1995 et les premiers sites internet de médias en France, médias traditionnels comme nouveaux médias en ligne n'ont toujours pas trouvé le « bon » modèle économique.

Un nombre croissant de médias met en place des « murs payants », Facebook les contourne en lançant « Instant Articles ». Les médias traditionnels investissent (enfin) dans le *live streaming*, ils sont concurrencés par Periscope ou Facebook Live. Les sites d'information à forte audience parviennent (un peu) à la monétiser grâce à des formes « innovantes » de publicité, un nombre croissant d'utilisateurs ont recours à des bloqueurs de publicité. Si l'on doit avoir une seule certitude, c'est qu'à l'heure où de nouvelles innovations technologiques apparaissent chaque jour, ce modèle économique devra être flexible.

La rapidité avec laquelle Internet évolue ne doit cependant pas nous empêcher d'en interroger les fondamentaux. Bien au contraire.

Ce travail de recherche inédit<sup>1</sup>, conjuguant les outils du *big data*, du *machine learning* et de l'économie, est basé sur la construction

---

<sup>1</sup> Ce livre est principalement basé sur les résultats d'une recherche scientifique qui a donné lieu à deux articles de recherche académique, technique et en anglais intitulés respectivement « The Production of Information in an Online World : Is Copy Right? » (Cagé, Hervé et Viaud, 2016) et « Estimating the Production and Demand for Online News : Micro-Level Evidence from the Universe of French News Media » (Cagé, Hervé et Viaud, 2017), et disponibles en ligne

et l'analyse d'une base de données unique en son genre : le contenu produit en ligne par les médias d'information en France sur une année (2013), combiné d'une part avec les caractéristiques humaines et économiques de ces médias – et en particulier la taille de leurs rédactions – et d'autre part avec des données extrêmement fines d'audience des différents sites, permettant d'analyser dans quelle mesure les citoyens adaptent leur consommation sur Internet à l'originalité des contenus produits.

*Une analyse inédite des contenus produits sur Internet*

Comment l'information est-elle produite et par qui ? Comment est-elle reprise et comment se diffuse-t-elle sur Internet ? Comment y est-elle consommée ?

Notre travail repose sur l'analyse des sites internet des médias d'information français, c'est-à-dire non seulement les sites internet des journaux, mais également ceux des radios, des chaînes de télévision et bien évidemment des *pure internet players*<sup>1</sup>. Il y a en effet aujourd'hui peu de différences « en ligne » entre ces différents médias, quel que soit leur support – écrit ou audiovisuel – « hors ligne ». Ce corpus est complété par l'intégralité des contenus (en français) de l'Agence France Presse (AFP).

À partir de l'étude systématique des documents textes produits et diffusés par ces médias sur leur site internet en 2013, nous avons développé un « détecteur d'événements » qui nous a permis de mettre à jour l'ensemble des événements traités par ces médias. En particulier, pour chaque événement, nous avons déterminé le *news breaker* – le média qui le couvre en premier – puis nous l'avons suivi en ligne seconde par seconde. Jamais la vitesse à laquelle un événement se propage sur Internet n'avait été étudiée de manière aussi précise.

Nous avons également développé un « détecteur de copie » afin de quantifier, pour chaque événement, le taux de copie entre chaque document et tous ceux qui le précèdent. Au-delà de la réactivité de

---

pour les lecteurs intéressés (voir <<https://sites.google.com/site/juliacagehomepage/research>>). Nous proposons ici une présentation plus accessible des principaux résultats de ces travaux de recherche. Nous remettons également ces résultats plus largement en contexte et discutons en détail leurs implications en termes de politique publique.

1 Les *pure internet players* ou *pure players* sont les médias qui n'existent qu'en ligne, comme *Mediapart* ou *Rue89* en France.

chaque média, nous évaluons ainsi son apport à la production d'une information originale.

Bien sûr, une grande partie du contenu produit par les médias ne se rattache à aucun événement particulier dans l'actualité, qu'il s'agisse d'informations très locales couvertes par la presse départementale ou régionale qui ne donneront pas lieu à des reprises par d'autres médias, en particulier au niveau national, ou de *slow journalism* ou « journalisme lent », des articles de long court qui ne sont pas en lien direct avec l'actualité. Cela peut être également le cas de longs entretiens avec des personnalités qui ne donnent lieu à aucune reprise, ou de tribunes et autres éditoriaux. Nous étudions également ces articles et nous mettons à jour une typologie des différents médias et de leurs choix rédactionnels<sup>1</sup>.

### Gratuité

Produire une information originale garantit-il aux médias qui le font une audience en ligne et des revenus suffisants? Si Internet a bousculé les médias traditionnels, c'est tout d'abord parce qu'il s'agit d'un formidable outil de diffusion – et de démocratisation – de l'information. Les audiences sur Internet sont, à première vue, sans commune mesure avec les ventes de la presse papier. Le nombre de visiteurs sur le site internet d'un média comme *Le Monde* varie chaque jour entre 1,5 et 2 millions. On est bien loin de la diffusion papier quotidienne du titre qui dépasse difficilement les 250 000 exemplaires en France, même si l'on augmente celle-ci du nombre de lecteurs par exemplaire<sup>2</sup>.

On pourrait discuter à l'infini des mérites respectifs des supports papier et digital. Tout le monde s'accordera au moins sur un point :

1 Bien sûr, nous sommes loin d'être les premiers à offrir une typologie des médias, on pense en particulier aux travaux de Gaye Tuchman et à son livre séminal *Making news. A study in the Construction of reality* (New-York, The Free Press, 1978). Dans une perspective différente, Daniel Hallin et Paolo Mancini proposent une typologie internationale des « systèmes médiatiques » et distinguent le modèle libéral, le modèle démocratique corporatiste et le modèle pluraliste polarisé. Voir Hallin, D. C. et Mancini, P. (2004). *Comparing Media Systems : Three Models of Media and Politics. Communication, Society and Politics*, Cambridge University Press; Hallin, D. C. et Mancini, P. (2012). *Comparing Media Systems Beyond the Western World. Communication, Society and Politics*, Cambridge University Press.

2 Toutefois, comme l'a montré l'un des auteurs de ce livre, le temps de lecture consacré à la lecture papier reste beaucoup plus important que le temps de lecture consacré à la lecture en ligne, où les actualités ne sont souvent que survolées. Voir Cagé, J. (2015). *Sauver les médias : Capitalisme, financement participatif et démocratie*, Paris, La République des idées/ Seuil.

le succès du support digital n'est pas étranger au fait qu'une partie de l'information y est disponible de façon gratuite<sup>1</sup>. Dès lors pourquoi payer?

D'autant que la propagation de l'information est extrêmement rapide sur Internet : il faut en moyenne moins de trois heures (175 minutes) pour qu'un événement couvert par un site d'information le soit également sur un autre site. La moitié des événements donne lieu à des reprises en seulement 25 minutes ; un quart des événements en seulement 230 secondes. Ces chiffres ne prennent en compte que les médias d'information eux-mêmes, pas les agrégateurs, ni les réseaux sociaux qui sont des facteurs supplémentaires d'accélération de la diffusion.

Cela n'est pas sans poser de problèmes pour la mise en place d'un modèle économique viable. Si un média, pour financer sa rédaction, décide de mettre en place un mur payant, c'est-à-dire de faire payer pour l'accès à ses contenus, mais que l'information qu'il produit est disponible en l'espace de cinq minutes sur le site d'un média concurrent mais lui gratuit, il y a fort à parier qu'un nombre important de citoyens choisira de ne pas payer et d'aller voir ailleurs.

### *Copié-collé*

D'autant plus qu'ailleurs, il y a un fort goût d'ici. On lit un peu partout que sur Internet le copié-collé a pris la place de l'« originalité ». Les sites des différents médias seraient substituables, et cette substituabilité serait grandissante avec l'essor des réseaux sociaux et des agrégateurs.

C'est bien le cas. Si les médias sont si réactifs, c'est en grande partie parce qu'ils « innovent » peu. Les médias – y compris les médias de référence – se contentent régulièrement de reprendre les contenus originaux produits par l'AFP ou leurs concurrents, sans même y ajouter d'efforts rédactionnels, voire en omettant parfois les procédures réglementaires de citations et de crédits. 64% de l'information produite en ligne par les médias d'information est du copié-collé pur et simple.

---

1 En tentant d'évaluer l'efficacité des murs payants ou *paywalls*, nombreux sont ceux qui comparent trop rapidement l'audience avant l'introduction du mur payant et le nombre d'abonnés à la suite de cette introduction. Or le nombre d'abonnés digitaux ne devrait pas être comparé à l'audience en ligne mais aux abonnés papiers avant l'introduction d'Internet.

*Fragilité ?*

Produire de l'information originale, c'est tout d'abord payer une rédaction et donner à ses journalistes les moyens de travailler. Cela suppose que la production d'information génère en contrepartie suffisamment de ressources pour se financer. La question clef réside donc dans la monétisation de la production d'information<sup>1</sup>.

Or le modèle économique des médias est rendu compliqué par son système de coût. La production d'information se caractérise en effet par des coûts fixes importants – en particulier les coûts relatifs à la taille de la rédaction – et des rendements d'échelle croissants<sup>2</sup>. Pour le dire autrement, le coût de production varie très peu avec le nombre d'exemplaires vendus ou l'audience, et ce plus encore à l'heure d'Internet où les coûts variables tendent vers zéro. Les coûts du papier, de l'impression ou de la distribution ont ainsi disparu dans le monde numérique.

Historiquement, les journaux ont été désireux de supporter ces coûts fixes élevés car en contrepartie, ils espéraient des profits. Publier une information en exclusivité pour un journal papier lui assurait, au moins durant une journée, une augmentation de ses ventes et donc de ses revenus, car cette information n'était pas disponible dans l'exemplaire papier des journaux concurrents<sup>3</sup>. Autrement dit, l'information avait alors une valeur commerciale.

Ce n'est plus le cas aujourd'hui. Sur Internet, sitôt publiée, l'information est disponible presque immédiatement sur le site des journaux concurrents. Comment monétiser dès lors la production d'une information originale ? Quelles incitations les médias ont-ils encore à payer pour une rédaction pléthorique alors qu'ils pourraient se contenter d'un petit nombre de journalistes en charge de monitorer le web et de faire du copié-collé, voire de simples algorithmes et de « journalistes robots » ? L'effondrement du nombre de journalistes de

1 À l'exception bien sûr de l'audiovisuel public, celui-ci étant financé en majeure partie par la redevance. L'audiovisuel public ne peut donc être approché d'un point de vue purement comptable en mettant face à face revenus et dépenses. Le niveau de « dépenses » de l'audiovisuel public est en effet un choix politique, fonction de la volonté citoyenne d'investissement dans la fourniture de ce bien public qu'est l'information. L'AFP, qui satisfait en partie une mission d'intérêt général, se situe de même en dehors de la logique purement comptable.

2 Cagé, J. (2014). « Media Competition, Information Provision and Political Participation », document de travail de l'Université d'Harvard.

3 Schudson, M. (1981). *Discovering the News : A Social History of American Newspapers*, Basic Books. Gentzkov, M. et Shapiro, J. M. (2008). « Competition and Truth in the Market for News », *Journal of Economic Perspectives*, 22 (2), 133-154.

presse quotidienne aux États-Unis – qui est passé de 55 000 en 2007 à moins de 35 000 en 2015 –, et de façon plus récente et mesurée (pour l'instant !) la diminution du nombre total de journalistes en France, témoignent en partie de la disparition des incitations financières liées à la vitesse de propagation de l'information sur Internet<sup>1</sup> (**FIGURES 1 & 2**). Si la tendance actuelle devait se poursuivre – et tout laisse à penser qu'elle pourrait malheureusement s'accélérer – alors on comptera tout juste 32 000 journalistes en France en 2025. Le nombre de journalistes de presse quotidienne aux États-Unis tombera, lui, sous la barre des 7 000. Et il ne faudra pas s'étonner de constater un appauvrissement de la qualité de l'information donnée à consommer aux citoyens.

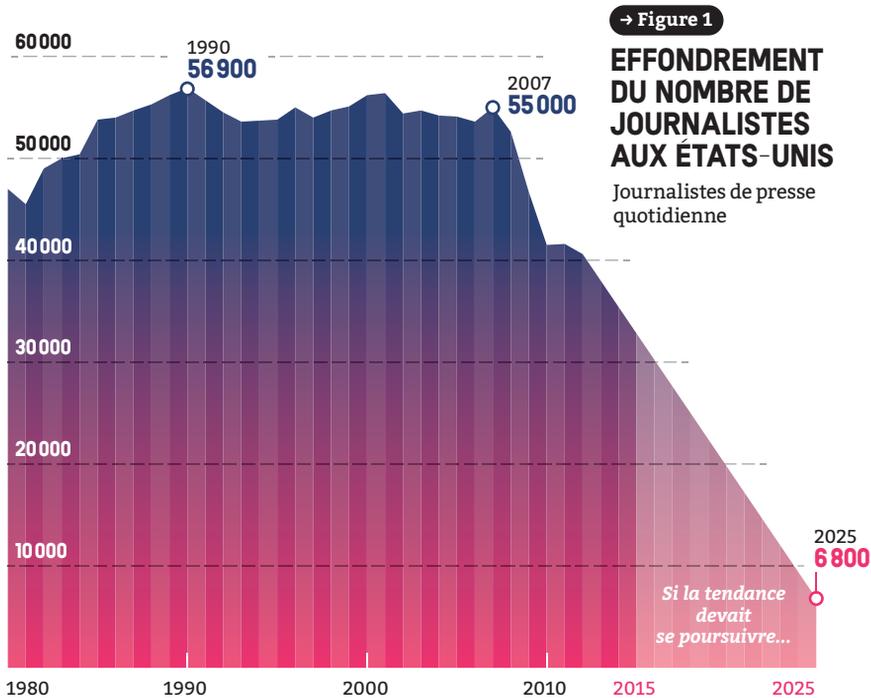
On observe d'ailleurs non seulement une baisse du nombre total de journalistes, mais également de la taille moyenne des rédactions comme l'illustre la **FIGURE 3** qui représente l'évolution de la taille moyenne de la rédaction de 25 titres de presse quotidienne régionale en France depuis 1980<sup>2</sup>. Le nombre moyen de journalistes pour ces titres est passé sous la barre des 50 depuis 2013. Aux États-Unis, le nombre moyen de journalistes par journal a été réduit par plus d'un tiers depuis 2001. Une telle réduction ne peut être sans conséquence sur la qualité et la quantité d'information originale produite.

D'après nos estimations, un journaliste supplémentaire dans une rédaction augmente en moyenne la production d'information *entièrement originale* de 28 articles par an<sup>3</sup>. De plus, une augmentation du nombre de journalistes conduit à une progression plus que proportionnelle de la quantité d'information originale produite. Autrement dit, l'émiettement des rédactions conduit à une recrudescence du copié-collé.

1 Bien sûr, la vitesse de propagation n'est pas la seule raison de cet effondrement. La chute vertigineuse des recettes publicitaires des médias traditionnels – à laquelle Internet n'est pas étrangère – explique également pourquoi les médias investissent moins dans leurs journalistes aujourd'hui. Voir en particulier Cagé, J. (2015). *Sauver les médias*, *op. cit.*, et Angelucci, Ch. et Cagé, J. (2016). « Newspapers in Times of Low Advertising Revenues », CEPR Discussion Paper 11414.

2 Les données utilisées ici sont tirées du travail de recherche historique de l'un des auteurs de ce livre : Cagé, J. (2016). « Payroll and Inequality within the Newsroom : Evidence from France, 1936- 2016 », document de travail, Sciences Po Paris. Cette recherche inédite sur l'évolution du nombre de journalistes en France et de leur rémunération témoigne également d'une précarisation croissante de la profession de journaliste.

3 Il s'agit ici de la production d'information originale classée à l'intérieur des « événements » ; les journalistes produisent également un certain nombre d'articles qui ne sont pas classés dans des « événements ». Nous aurons l'occasion d'y revenir.

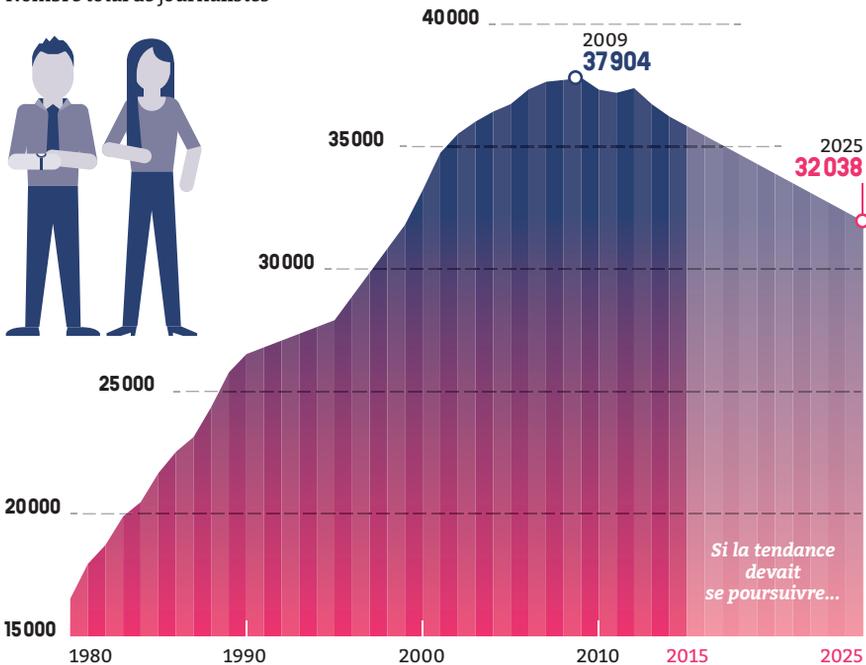


Source des données : ASNE annual employment census.

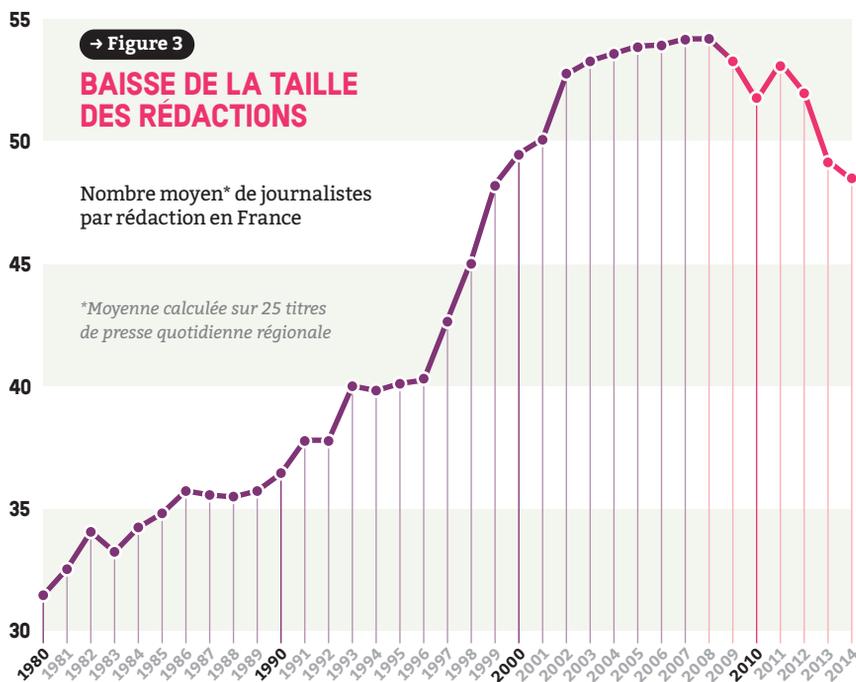
→ **Figure 2**

## BAISSE DU NOMBRE DE JOURNALISTES EN FRANCE

Nombre total de journalistes



Source des données : Commission de la carte d'identité des journalistes professionnels (CCIJP).



Source des données : calculs effectués par Cagé, J. (2016) à partir des données de la CCIJP.

La taille moyenne des rédactions des médias d'information en France diminue de 1% par an depuis 2013. Or 1% de journalistes en moins, c'est 1,20% de moins d'information *originale* produite. Toutes choses égales par ailleurs, si la diminution de la taille moyenne des rédactions devait continuer au rythme actuel, cela impliquerait qu'en 2020, ce seraient plus de 30 000 articles d'information entièrement originaux *de moins* qui seraient produits qu'en 2013.

Or, au vu des changements actuels, 1% de baisse par an est une estimation basse de la diminution de la taille des rédactions<sup>1</sup>. Un journal comme *L'Express* est passé de plus de 120 cartes de presse en 2013 à 40 aujourd'hui, soit un retour à la situation du journal... en 1965. *L'Obs*, qui a annoncé en juin 2016 un nouveau plan de départs volontaires, avait déjà vu le nombre de ses cartes de presse diminuer de 50 entre 2014 et 2015. Et c'est sans parler d'I-télé qui en 2016 a perdu plus de la moitié de sa rédaction suite au bras de fer qui a opposé les journalistes à un actionnaire intrusif et peu soucieux de déontologie, Vincent Bolloré pour ne pas le nommer. Certes, l'approche quantita-

1 Et seule la croissance des nouveaux *pure players* vient compenser l'écrémage des rédactions des médias traditionnels.

tive présente de nombreuses limites mais les tendances actuelles font peser une véritable menace sur l'information originale<sup>1</sup>.

### *Citations et réputation*

Serions-nous entrés dans un cercle vicieux irréversible où chaque innovation apportée par Internet et permettant une propagation de l'information encore plus rapide et systématique réduirait un peu plus les incitations des médias à investir dans la production d'une information originale? Quelles sont les solutions qui peuvent permettre d'échapper à cet engrenage?

Pour répondre à cette question, il faut tout d'abord évaluer dans quelle mesure les citoyens « récompensent » les médias pour la production d'une information originale. Ce n'est pas parce qu'une information est disponible en l'espace de quelques minutes « partout » que les consommateurs n'ont pas des préférences quant à l'endroit où ils souhaitent la lire; et il est possible que ces préférences se portent sur le média ayant le premier mis à jour la nouvelle. Certes, près de la moitié des utilisateurs arrivent aujourd'hui sur le site internet des médias « indirectement », en cliquant sur les liens qui leur sont proposés par les agrégateurs ou les réseaux sociaux. Mais cela signifie également que la moitié des utilisateurs continuent quant à eux à se rendre directement sur le site de leur(s) média(s) préféré(s). Et c'est là où l'effet marque – ou réputation – peut jouer un rôle clef<sup>2</sup>. Les citoyens récompensent-ils au final les « bons » médias?

Si l'audience agrégée totale des sites d'information augmente les jours où des événements « importants » surviennent (comme on pouvait s'y attendre), cette augmentation ne profite en moyenne que très faiblement aux médias que l'on pourrait qualifier d'« innovants », c'est-à-dire au *news breaker* et à ceux qui contribuent à la production d'information originale sur l'événement. Leur part d'audience augmente certes de façon statistiquement significative consécutivement

1 Il s'agit bien ici d'information *originale*. La quantité *totale* d'information disponible n'est pas près de diminuer. Mais l'information produite est de plus en plus homogène.

2 La rapidité avec laquelle un média traite une information influence également son rang – et donc sa visibilité – sur les sites internet des agrégateurs. Même si les agrégateurs ne sont pas au cœur de notre analyse – puisque nous n'étudions pas le contenu disponible sur leur site – ils jouent néanmoins indirectement un rôle crucial. En particulier, ils expliquent pourquoi un nombre important de médias tend à publier plusieurs fois le même article, en ne modifiant que le titre, ou à le publier sur plusieurs flux RSS, voire à modifier artificiellement l'heure de sa publication.

à la production de cette information originale, mais la magnitude de cette augmentation est extrêmement faible, voire économiquement insignifiante.

Cela ne signifie pas pour autant que les médias n'ont plus aucune incitation (économique) à produire de l'information originale. Si, au jour le jour, les variations de parts de marché ne viennent que faiblement récompenser les médias les plus innovants, au niveau agrégé, la production de contenu original par un média est fortement corrélée à l'audience moyenne de son site internet. En moyenne sur 2013, augmenter chaque jour d'un article la production d'information originale augmente le nombre moyen de visiteurs uniques quotidiens de 14000, ce qui est loin d'être insignifiant pour un grand nombre de médias. Autrement dit, la production de contenu original semble entraîner sur le moyen terme un effet de réputation qui se répercute sur les audiences en ligne. C'est cet effet de réputation qui expliquerait pourquoi les médias sont toujours incités à produire une information de qualité<sup>1</sup>.

De plus, il est important de noter que cet effet de réputation ne joue pas seulement « en ligne », mais également « hors ligne ». Au niveau mensuel, la production d'une information originale en ligne est en effet corrélée, pour les journaux papiers, avec les ventes en kiosque<sup>2</sup>.

L'importance de l'effet de réputation est renforcée par le rôle clef joué par les « citations », autrement dit, la mention d'un média comme source de l'information. Au niveau quotidien, le principal déterminant causal de la part d'audience d'un média est en effet le nombre de fois où ses concurrents y font référence<sup>3</sup>. Cela peut être dû à l'utilisation de liens hypertextes et/ou à la volonté des utilisateurs de revenir « à la source ».

D'où l'importance des crédits adéquats. Or, les médias tendent *de facto* à « oublier » le plus souvent de citer les médias qu'ils copient : ils ne le font que dans 8% des cas. Même quand l'AFP n'est pas le *news breaker*, les médias préfèrent souvent faire référence à l'agence

1 Ces résultats empiriques sont consistants avec les prédictions théoriques de Gentzkow, M. et Shapiro, J. M. (2006).

2 Même si bien sûr cela donne lieu à des problèmes intrinsèques de mesure, la production de contenu original en ligne s'accompagnant le plus souvent également de la production de contenu original sur la version papier.

3 Nous définissons ici comme « citation » ou « référence » la mention du nom du média comme source de l'information, par exemple « comme l'a révélé le journal *Le Monde* ».

plutôt qu'au média qui a originellement sorti l'information. Plus de la moitié des citations sont des références à l'AFP. Et dans la majorité des articles publiés en ligne, on ne trouve pas une seule référence à un média concurrent<sup>1</sup>.

*Pour une régulation des médias adaptée à Internet*

Le régulateur – en France mais également ailleurs dans le monde – ne semble pas avoir pris la pleine mesure du changement de paradigme qui caractérise les médias au <sup>xxi</sup>e siècle, à l'âge d'Internet<sup>2</sup>.

D'une part, la régulation française des médias se fait encore principalement par support. Alors que le Conseil Supérieur de l'Audiovisuel (CSA), comme son nom l'indique, est en charge de la communication « audiovisuelle », c'est-à-dire de la télévision et de la radio<sup>3</sup>, c'est la Commission Paritaire des Publications et Agences de Presse (CPPAP) qui régule les publications papiers ainsi que les services presse en ligne. Pourtant, lorsque l'on compare leur production d'information sur Internet, le support « hors ligne » importe peu. Pourquoi dès lors ne pas adopter une approche véritablement « trans-media » de la régulation ? Nous faisons des propositions en ce sens à la fin de ce livre.

D'autre part et surtout, nous pensons que l'accent a été insuffisamment mis jusqu'à présent sur les questions de réputation et sur la problématique pourtant clef de la propriété intellectuelle, des citations et des crédits.

Jusqu'à présent, la question du copyright a été abordée principalement du point de vue de la relation entre médias et agrégateurs. Mais cette question est en fait beaucoup plus large et le respect de la propriété intellectuelle doit être pensé également dans les relations entre médias producteurs d'information originale.

Ce qui est essentiel aujourd'hui, c'est de s'assurer que les médias conservent des incitations suffisantes pour produire une information originale et de qualité, et continuent de financer le travail de

1 Ce qui peut sembler étonnant puisque – on l'a dit – plus de la moitié du contenu produit en ligne est du copié-collé.

2 D'ailleurs, il est frappant de noter qu'en France, une régulation aussi importante que celle qui encadre la concentration des médias date des années 1980, bien avant l'invention d'Internet ! Or, aucune force politique ne semble faire de cette régulation une priorité, dans un contexte de concentration croissante du secteur.

3 Ainsi que, depuis 2004, des « services interactifs » – autrement dit des télévisions et des radios en ligne – et des services de médias audiovisuels à la demande depuis 2009.

leurs journalistes. En l'absence de telles incitations et alors même qu'Internet est un formidable outil de démocratisation et de production de l'information, alors même que les médias ont gagné au cours des vingt dernières années tout à la fois en termes de réactivité et en termes de visibilité, nous risquons – en tant que citoyens – d'être moins informés demain que nous ne le sommes aujourd'hui.

Cela est inquiétant pour le futur de nos démocraties. Si nous avons fait le choix dans ce livre de nous focaliser sur les médias d'information, c'est parce que nous sommes convaincus que l'information que produisent ces médias est un bien public. Ce bien public, il est nécessaire aujourd'hui de le repenser pour le sauvegarder. Nous sommes également convaincus qu'une meilleure régulation permettrait de tirer toutes les potentialités d'Internet et de contribuer à un débat démocratique mieux informé.

### *Limites de l'analyse*

Les réponses que nous apportons dans ce livre sont bien sûr loin d'être définitives. Les limites de notre analyse sont nombreuses. Les données que nous exploitons proviennent d'un seul pays – la France – et d'une seule année – 2013 – qui peut sembler une éternité tant les choses changent rapidement à l'heure d'Internet. Ainsi en 2013, dans notre base de données, très peu de médias en France se trouvent derrière un mur payant (*paywall*), alors qu'ils sont aujourd'hui de plus en plus nombreux à avoir fait ce choix. Nous ne pouvons donc fournir que des premières pistes de réflexion concernant l'impact de ces *paywalls* sur la production d'information sur Internet. Nos futurs travaux couvriront une période et un périmètre plus larges, permettant d'aborder ces questions.

Autre limite de taille, notre approche est purement quantitative : elle permet une analyse systématique et la mise à jour de causalités. Le recours à des algorithmes et à des régressions économétriques a toutefois ses propres faiblesses au regard d'une approche plus qualitative qui, pour chaque événement pris individuellement, permettrait sans doute d'apporter un certain nombre de conclusions certes moins systématiques, mais plus riches et détaillées.

Toutefois ce travail, fruit de plusieurs années de recherche et d'une combinaison inédite des outils des sciences informatiques et de l'économie, a le mérite d'éclairer certains aspects d'une problématique extrêmement complexe. Il pourra ainsi, nous l'espérons, ins-



## INTRODUCTION

pirer la mise en œuvre de politiques publiques plus pertinentes. En ouvrant ici la boîte noire de la production d'information sur Internet, notre objectif premier est d'ouvrir le débat bien au-delà du seul champ académique. Nous espérons que nombreux sont ceux qui viendront y prendre part, citoyens, journalistes comme décideurs publics !



## CHAPITRE 1

### L'INFORMATION SUR INTERNET : UNE APPROCHE TRANSMEDIA

#### Une base de données unique

Notre corpus est composé du contenu produit en ligne par les médias d'information politique et générale en France en 2013. Plus précisément il est constitué de 86 médias d'actualité : l'Agence France Presse (AFP), 59 journaux (35 titres de presse quotidienne régionale, 7 titres de presse quotidienne nationale, 3 gratuits, 12 hebdomadaires nationaux et 2 mensuels), 9 télévisions, 7 radios et 10 *pure players* (produisant uniquement de l'information en ligne) (**FIGURE 4**<sup>1</sup>). Pour chacun de ces médias, nous avons collecté les articles publiés en ligne en 2013 (que ces articles soient disponibles à tous ou derrière un mur payant) en utilisant leurs flux RSS principaux et/ou leurs Sitemaps<sup>2</sup>. Ce corpus est largement issu du projet de recherche OTMedia dont l'INA était le pilote (**ENCADRÉ**). Il a été complété *a posteriori* par de nouvelles captures pour les médias qui n'étaient pas couverts dans le cadre d'OTMedia.

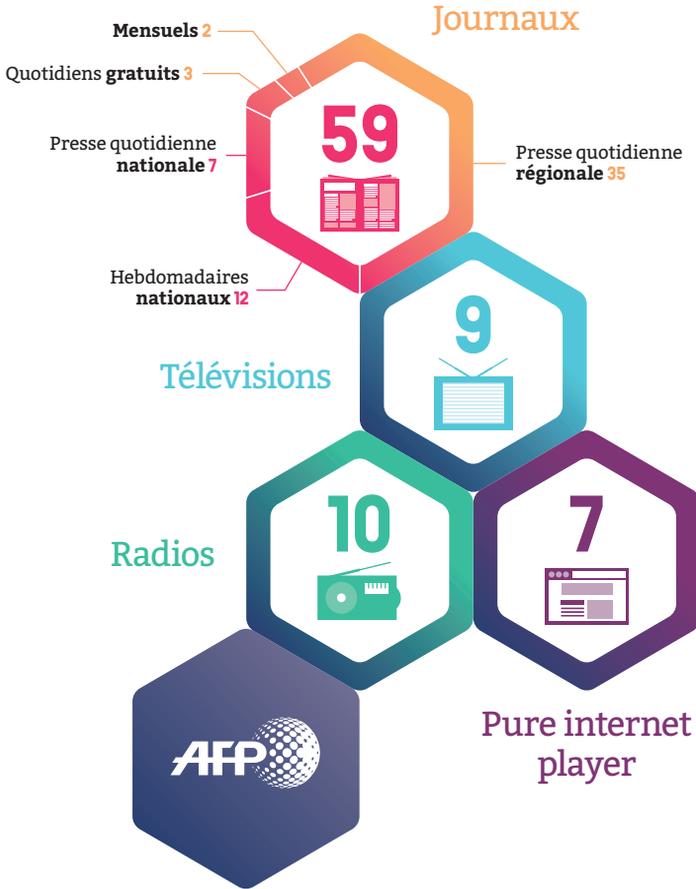
---

1 Un petit nombre de médias qui avaient pourtant le label Information Politique et Générale (IPG) en 2013 en France sont absents de notre corpus. La majorité d'entre eux n'ont tout simplement pas de site internet (c'était le cas de nombreux titres de presse hebdomadaire départementale et régionale en France en 2013). Pour les autres, nous avons malheureusement été dans l'incapacité technique de collecter leurs contenus.

2 À l'exception de l'AFP dont l'INA a reçu directement l'intégralité des dépêches.

→ Figure 4

# UN CORPUS COMPOSÉ DE 86 MÉDIAS D'ACTUALITÉ



## **Le projet OTMedia**

Le projet de l'Observatoire TransMedia (OTMedia) a été subventionné par le programme Contint de l'Agence Nationale de la Recherche (ANR) de 2010 à 2014. Il a réuni des partenaires technologiques et en sciences humaines et sociales : l'INA – coordinateur du projet –, l'INRIA, l'AFP, la société Syllabs, le Laboratoire Technique, Territoires et Société de l'École des Mines, le Laboratoire d'Informatique d'Avignon, et le Laboratoire Communication, Information, Médias de l'Université Sorbonne Nouvelle - Paris 3.

L'objet de cet Observatoire est l'étude et la traçabilité des événements médiatiques sur tous les supports de diffusion : web (blogs, sites des médias traditionnels...), presse, radio et télévision. Une plateforme d'analyse des différents flux médiatiques, quelle que soit leur modalité (texte, audio, vidéo, images), a été conçue pour détecter, circonscrire, suivre, mesurer, analyser et étudier la propagation des événements médiatiques, les thématiques émergentes, les sujets de « buzz », ainsi que les évolutions de vocabulaire.

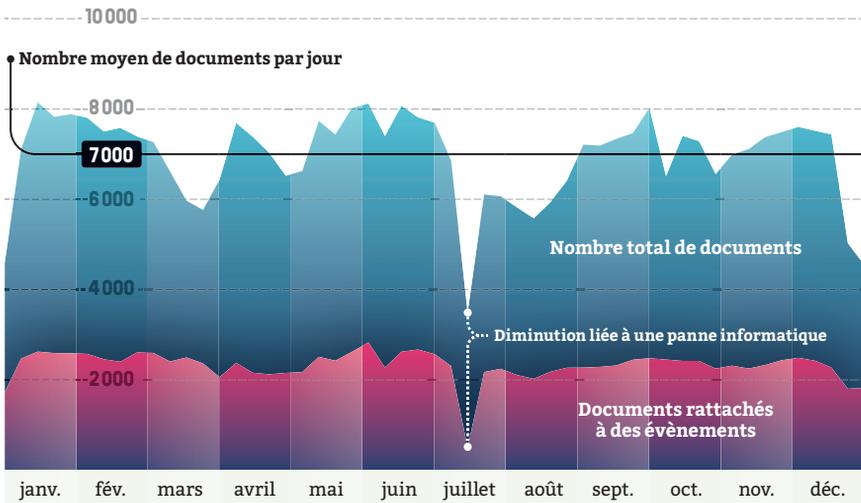
Cette plateforme fait aujourd'hui l'objet d'un développement interne à l'INA. Elle met en œuvre des modules de collecte, d'analyse des modalités (segmentation, transcription, description d'images et de vidéo, analyse de texte) et de fouille de données avancées (détection de copies, de référencement, d'entités nommées, d'événements) sur des volumes d'information conséquents (plus de dix millions d'objets).

L'ambition de cette plateforme est de donner aux chercheurs en sciences humaines et sociales et aux acteurs de l'information les moyens d'analyser et de comprendre l'information aujourd'hui. Les résultats présentés dans ce livre sont un exemple représentatif des objectifs de ce projet.

→ Figure 5

## UNE BASE DE DONNÉES DE PLUS DE 2,5 MILLIONS DE DOCUMENTS DONT 1/3 CLASSÉS DANS DES ÉVÉNEMENTS

Nombre de documents par jour

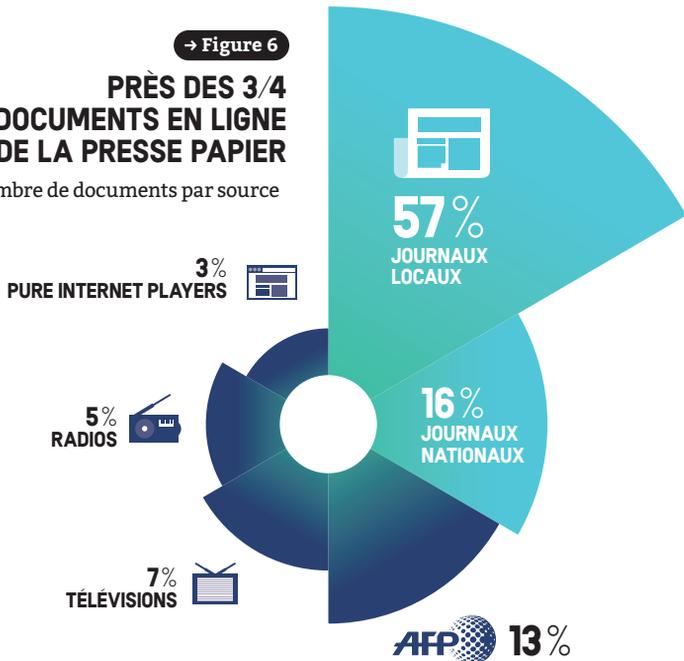


Un événement ici est un événement médiatique. Les événements sont définis à partir de notre algorithme de détection d'événements, décrit en détails dans le texte, qui nous a permis de mettre ensemble tous les articles se rapportant au même « fait ».

→ Figure 6

## PRÈS DES 3/4 DES DOCUMENTS EN LIGNE ISSUS DE LA PRESSE PAPIER

Nombre de documents par source



Notre corpus comprend 2 548 634 documents pour l'année 2013, soit en moyenne environ 7 000 documents par jour<sup>1</sup> (FIGURE 5). La taille moyenne de ces documents est de 1 865 caractères.

73% de ces documents proviennent des sites internet de la presse papier (57% des journaux locaux et près de 16% des journaux nationaux), 6,6% de sites internet de chaînes de télévision, 4,7% de sites internet de stations de radio, 12,5% de l'AFP et 3,2% des *pure players* (FIGURE 6).

## Une approche transmedia

L'approche transmedia est l'une des dimensions clefs de ce travail de recherche. Il est devenu presque impossible de reconnaître à partir de son seul site web le support originel – hors ligne – d'un média (presse écrite, radio, télévision...). Cela apparaît très clairement sur la FIGURE 7 qui compare quatre captures d'écran de sites internet de médias dont le support originel n'est pas le même (*Le Monde*, France Télévision, Europe1 et *Slate*).

Sur Internet, les médias utilisent souvent les mêmes logiciels, les mêmes techniques, et le travail d'un journaliste est relativement similaire sur le site d'une télévision ou sur celui d'un journal papier. Tous les médias ont recours en ligne au texte, à la vidéo et à la photo<sup>2</sup>.

Il nous a ainsi paru essentiel de mettre sur le même plan tous les producteurs d'information sur Internet, quel que soit leur statut « hors ligne<sup>3</sup> ». Et cela semble d'autant plus pertinent que le recours au copié-collé, ou au contraire la production d'information originale,

1 Pour certains jours, notamment au mois de juillet 2013, le nombre de documents à l'intérieur de notre corpus est plus faible qu'en moyenne. Cela provient à la fois de la moindre production d'articles par les médias durant les mois d'été – ainsi que durant les weekends – mais également de problèmes techniques de capture liés aux aléas de l'informatique. Une panne de serveur un samedi à l'INA ne donnant logiquement lieu à une réparation que le lundi suivant peut ainsi expliquer une chute dans le nombre de documents capturés.

2 Voir par exemple, à ce sujet : Peitz, M. et Reisinger, M. « Chapter 10 - The Economics of Internet Media » in Waldfogel, J., Anderson, S. P. et Strömberg, D. eds. (2016). *Handbook of Media Economics*, Volume 1, North-Holland, pp. 445-530. Nous nous limitons cependant ici à l'analyse des documents texte. L'étude de la photo et de la vidéo sera l'objet de futures recherches.

3 Ce qui ne veut pas dire que nous ne « contrôlons » pas, dans l'analyse empirique, pour le support hors ligne des médias (avec l'introduction de variables indicatives). Mais tous ces médias sont inclus dans l'analyse indépendamment de leur support hors ligne.

→ Figure 7

### Policiers : « Réduire le problème à des questions d'effectifs, c'est jeter de l'huile sur le feu »

ENTRETIEN



Des policiers manifestent dans toute la France pour exprimer leur malaise. Le sociologue Sébastien Roché analyse cette colère qui ne faiblit pas.

- Le gouvernement ne peut pas se permettre un conflit avec la police
- On est tellement en colère qu'on continuera tous les soirs
- L'exécuff tente d'endiguer le « Nuit debout des policiers »
- Policiers en colère : François Hollande va recevoir les syndicats à l'Élysée

### A NE PAS MANQUER

L'entreprise Pump (25) : postures et impostures

ÉPIGRAMME  
Nos écrans nous coupent-ils la parole ?

ENTRETIEN  
« L'Etat de droit reste le fondement de l'Union européenne »

Alap, sans médecins ni chirurgiens

REPORTAGE  
Sylviane Noillard-Roux, 79 ans #EnMémoire

REPORTAGE MONDIAL  
Statistiques ethniques : soyons réalistes !

Tous les éditeurs

### EN IMAGES



Mossoul, aux premières heures de la bataille décisive

Le Monde  
Faites plaisir !  
Effectuez un abonnement

### La défaite de Duflot, sans incidence à gauche

Les prétendants à la primaire du Parti socialiste se disent prêts à travailler avec le futur candidat vert.

### Au secours, mon sex-toy m'espionne !

En anses, le monde compte vingt cinq milliards d'objets connectés, autant de cibles potentielles pour les pirates informatiques.

### Le Monde

Le journal date du 22 octobre

17:22 Rapport de la base sur le charbon

17:18 « Pokémon Go » à l'étranger

17:14 Le récit d'une ex-condamnée à mort

17:14 TV : Les Hommes de l'ombre

17:10 TV : On refait le boulevard

17:08 Nintendo 3DS : un en et deux de formats

17:03 Migrants : L'exilar violentement critiqué

16:58 CETA, l'écrit avec Wallons et Canadiens

16:50 Un campement rom devant le Sénat Open

16:47 Militaires : une mobilisation inédite

16:46 L'acteur Clément Michu est mort

16:38 Les qualités de l'ingénieur du futur

16:43 Grand banditisme : prison de 30 ans

16:42 Un nouveau leader mondial du tabac

16:37 Le Monde est-il espionné ?

16:42 L'audience du Tour de France

▲ Le Monde



PARENTALITÉ  
Cet emoji d'une femme qui allaite d'est pas une bonne idée



MACROECONOMIE  
Elise Lucet VS Estrosi: le travail de démolition des journalistes a commencé



MACROECONOMIE  
Le consultant Macron ne pourra pas sauver la France



MADAME PROPRE  
Qu'est-ce qu'un soutif sale ?



MYSTÈRES D'ÉCRIVAIN  
Le petit secret de R.L. Stine, le père de la série « Chair de Poule »

EN BRÈVE

**OSCAROUBIE**  
Votre idée sur Kickstarter se peut être déjà en vente... en Chine

**EROKH 500**  
Donald Trump a-t-il aussi un gros problème avec ses mails

**CRIME EN LANGUE**  
Les liens entre la mafia italienne et l'État islamique dévoilés par un journaliste

**DISCRIMINATIONS**  
L'homophobie nuit gravement aux résultats scolaires

**TASTY**  
Candy Crush va devenir un jeu télé

**SANTÉ**  
La pilule est-elle historiquement soignée ?

**ARCHÉOLOGIE**  
Des chercheurs ont peut-être retrouvé la première victime d'un boomming

**MIAMI**  
Pourquoi vous aimez manger des cookies avec votre café

PLUS DE EN BRÈVE

**Slate** PRODUITS

**Titane, Nuts & les sales grosses #5**  
Combien plus un secret de famille ?

**Transfert #9**  
Un jour, Nathalie est devenue amie avec Ulrich, désormais elle ne fera plus confiance à personne

▲ Slate



→ Figure 7

**LE LIVE**

Météo Vidéo

POLICIERS #FOOT #VILLE #PROGRAMME #VELLY

---

**Franceinfo**

Après les violences dans un lycée de Tremblay-en-France (Seine-Saint-Denis), les profs se disent "choqués et inquiets". Mais pas résignés : après l'expulsion d'une professeure du lycée Natoire-Rocher, Anne Rioux, pour harcèlement, a recruté [jeunes.lesfranciscains.com](#)



(Capture d'écran Google Street View)

---

**Franceinfo**

En Allemagne, un policier a été tué et trois autres blessés par un missile des "kayakers de Rasch", une mouvance néofasciste et de droite à l'extrême droite, rapporte [Le Monde](#).

---

**Elodie Drouard** Franceinfo

Pour la deuxième année consécutive, le Comité Con Paris, version française de la plus grosse fête de la plus célèbre d'Internet, vendredi 21 octobre dans la capitale. L'occasion pour le réalisateur d'y exhiber son plus beau chapeau et d'y croquer, entre deux conférences, quelques savoureux aperçus sans sans sans... de comités. L'excuse parfaite trouvée par cette fête d'aller pour vous présenter leurs coups de cœur comités de ces dernières semaines.



### Primaire à droite : pourquoi la campagne de Bruno Le Maire ne décolle pas

→ Primaire à droite : Juppé maintient son avance sur Sarkozy, selon un sondage Harris Interactive pour France Télévisions  
 → VIDEO, Recyclée, en vogue, ou nouvelle : trois propositions de Bruno Le Maire

---

**MONDIALE**



### Après les violences dans un lycée de Tremblay-en-France, les profs "choqués et inquiets" mais pas résignés

▲ France Télévision

POLITIQUE INTERNATIONAL SPORT MEDIAS CULTURE ECONOMIE SOCIÉTÉ FAITS DIVERS SANTÉ



### Policiers : pourquoi la contestation monte d'un cran

Depuis deux jours, des centaines de policiers de la piste organisent des rassemblements spontanés pour réprimer leur "ton le-bât", partout en France.

**Combattifs démontre la "patte" du FN dans les manifestations de policiers**  
**Bernard Cazeneuve se rassure les significatifs de policiers en colère**  
**Marseille : manifestation surprise de policiers sur le Vieux-Port**

---

**"Individu de race noire" : les coulisses d'une polémique**  
 Un clip de la chaîne "rap de race noire" lors de la diffusion mardi soir d'une alerte enlèvement a suscité un tollé.

---

**Djamel : retrouvée saine et sauve : les circonstances de son enlèvement**  
 Le père de Djamel, qui avait enlevé sa fille de quatre mois, mardi à Grenoble, s'est rendu mercredi en fin de semaine. L'enfant est en bonne santé.

• **Alerte enlèvement : le père s'est rendu, le nourrisson retrouvé en bonne santé**  
 • **"Individu de race noire" : quand l'alerte enlèvement vire à la polémique**

**LE FLASH ACTU**

- 17329 • La cité d'Ivry d'un ministre belge après le piratage de son compte Twitter
- 17324 • Visibilité de la Foucault : pas de bouillottes prévues ce week-end
- 17319 • Google veut nous faire payer moins cher nos mails d'archive
- 17312 • Pascal Clémex about un éléphant et crée la polémique
- 17305 • Serie la "Spice humanitaire" jouit à Abaj dans 15 heures
- 16348 • La météo de jeudi : instable dans le nord et le nord-est, ensoleillée dans le sud
- 16337 • Accident industriel en Allemagne : le bilan monte à trois morts
- 16326 • Trois jeunes interpellés pour avoir menacé d'incendier un lycée

▲ Europe 1



ne varie pas au final avec le support hors ligne, mais bien avec la taille des rédactions<sup>1</sup>. Pas plus d'ailleurs que les audiences sur Internet.

Dans une analyse centrée sur les *breaking news*, il était également indispensable d'inclure l'AFP, principale agence de presse française. Certes, cette dernière ne s'adresse pas directement au grand public mais à ses clients et elle diffère à bien des points de vue des autres médias. Mais il est essentiel de la prendre en compte dans l'étude de la production de l'information : ne pas le faire aurait conduit à faussement attribuer la paternité d'événements à des médias qui reproduisent en fait une information de l'AFP.

Bien sûr, des différences fondamentales demeurent entre les différents types de médias, et radios et télévisions sont aujourd'hui moins dépendantes des revenus publicitaires qu'elles tirent de leur site internet que les journaux. Radios et télévisions font également moins directement concurrence à leur support hors ligne par le contenu qu'elles mettent à disposition – le plus souvent gratuitement – sur leur site internet, car ces contenus (textes) sont moins substituables<sup>2</sup>. Mais ces différences sont vraisemblablement appelées à s'effacer de plus en plus dans le futur.

Alors que pendant des décennies on pouvait distinguer d'un côté les journaux, financés tout à la fois par les recettes de ventes et d'abonnement et les recettes publicitaires, et de l'autre les radios et les télévisions, financées uniquement par la publicité<sup>3</sup>, depuis le développement de la télévision par câble aux États-Unis dans les années 1970, de plus en plus de chaînes – comme HBO et Showtime – sont financées uniquement par abonnement et ne diffusent plus de publicité. La radio par satellite, apparue aux États-Unis en 2001, se finance de même uniquement par abonnement, alors que la radio sur Internet est financée par une combinaison d'abonnement et de publi-

1 Insister sur une approche transmedia est nécessaire. Dans le cadre du lancement de la version française de Mashable par France 24, on a pu entendre qu'il n'était pas du domaine de l'audiovisuel public de venir concurrencer en ligne des médias privés. Or, le site internet de France Télévisions se range régulièrement parmi les cinq sites d'information les plus consultés.

2 Même s'il est à noter que la consommation de radio et de télévision sur Internet est en développement croissant, notamment avec l'utilisation des podcasts qui favorisent la délinéarisation de la consommation.

3 Nous parlons ici des radios et des télévisions privées, le financement de l'audiovisuel public reposant également en France sur la redevance qui existe sous différentes formes dans les autres pays.

cité<sup>1</sup>. Et ce phénomène risque de s'accélérer, alors que les recettes publicitaires des médias traditionnels ne cessent de s'effondrer.

## Une information... en continu ?

Pour chacun des 2,5 millions de documents inclus dans notre échantillon, nous connaissons l'heure de publication fournie par le média. Quelles sont les stratégies de publication des différents acteurs ? La **FIGURE 8** représente l'ensemble des documents dans notre base de données selon leur heure de publication (par minute). Plusieurs enseignements peuvent en être tirés.

Tout d'abord, au-delà de l'idée largement répandue d'une « information en continu », il apparaît clairement que le cycle de publication de l'information en ligne en France est essentiellement un cycle de 12 heures, de 6h00 du matin à 18h00<sup>2</sup>. Malgré la révolution numérique, ce cycle ressemble étrangement à celui décrit par Gaye Tuchman il y a près de quarante ans dans le cadre de son étude des conversations à l'intérieur des rédactions : quelques journalistes arrivent au bureau dès 8h00 du matin – au cas où il arriverait quelque chose –, la plupart vers 10h00, et tous partent entre 18 et 19h00, après que leurs « sources » potentielles sont elles-mêmes rentrées chez elles<sup>3</sup>.

Ce cycle se caractérise également par une diminution de la quantité d'informations mises en ligne à l'heure de la pause déjeuner, entre midi et 14h00. On pourrait interroger la pertinence d'une telle stratégie de publication mais elle correspond en fait au comportement de lecture de la plupart des utilisateurs d'Internet, qui consomment massivement l'information pendant leurs heures de bureau (ce que Pablo Boczkowski a appelé « *news at work* »).

1 Voir par exemple Anderson, S. et Waldfogel, J. (2016). « Preference Externalities in Media Markets », CEPR Discussion Paper 10835.

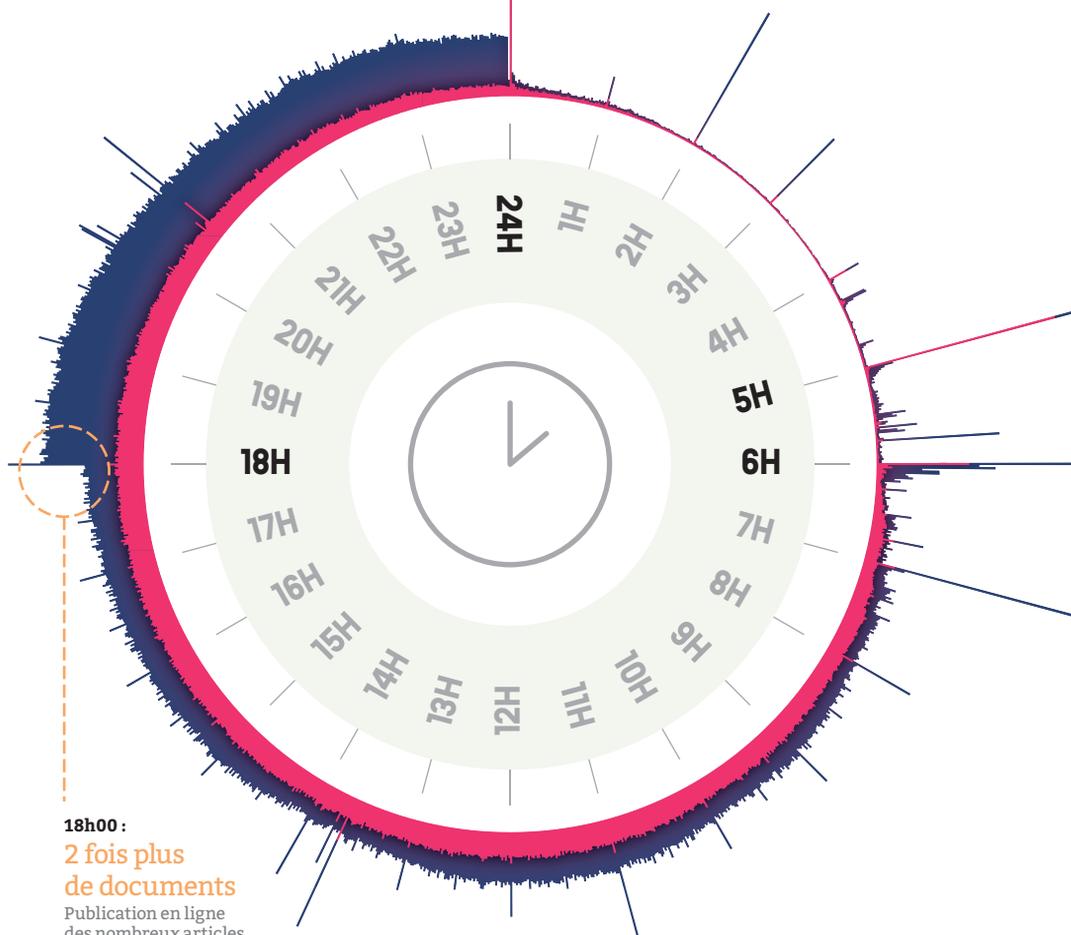
2 Cela est peut-être appelé à changer dans le futur. Aux États-Unis, un journal comme le *Washington Post* a massivement investi au cours des dernières années dans la production d'information en continu avec la mise en place d'une rédaction « de nuit » en charge de publier des articles et de rafraîchir le site internet du journal en dehors des heures « traditionnelles » de publication. Le *Wall Street Journal* a de même mis en place en 2014 un service d'informations en temps réel (*real-time news desk*) avec un rédacteur en chef en charge de la nuit. En France, un certain nombre de publications de référence ont également pris en compte récemment la nécessité de rafraîchir leur site internet à intervalles réguliers dans le choix de la localisation de leurs bureaux à l'étranger.

3 Tuchman, G. (1978). *Making news. op. cit.*

→ **Figure 8**

# 24 HEURES DE PUBLICATION

Documents **classés** et **non classés** dans des événements par horaire de publication (total sur l'année)



00h00 :

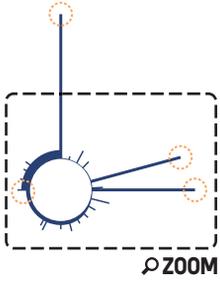
**59 000 documents**

Pic quotidien le plus important : à minuit, *L'Indépendant*, *Sud Ouest* et *Le Télégramme* publient automatiquement en ligne les articles à paraître dans leur édition papier

18h00 :

**2 fois plus de documents**

Publication en ligne des nombreux articles des multiples éditions du journal *Ouest France*



**05h00 :**

**38 000 documents**

Publication automatique en ligne  
des articles des journaux du groupe EBRA

**06h00 :**

**42 000 documents**

Publication automatique en ligne  
des articles du journal *Le Dauphiné Libéré*

La mise en ligne d'informations est pourtant loin d'être uniquement « humaine ». En particulier, les journaux quotidiens régionaux du groupe EBRA (*Le Bien Public*, *Les Dernières Nouvelles d'Alsace*, *Le Journal de Saône et Loire*, *L'Est Républicain*, *Le Républicain Lorrain*, et *Vosges Matin*) semblent avoir automatisé la publication en ligne des articles de leur édition papier du jour à 5h00 du matin. Ce qui témoigne également de l'uniformisation des stratégies de publication de rédactions rendues exsangues par un actionnaire – le Crédit Mutuel – en position de monopole sur tout l'Est de la France. De même, *L'Indépendant* et *Sud Ouest*, tous deux propriété du groupe Sud Ouest en 2013 (*L'Indépendant* est passé depuis 2015 sous le contrôle du Groupe La Dépêche) publient massivement ces articles à minuit, tout comme *Le Télégramme*, alors que *Le Dauphiné Libéré* semble lui avoir choisi de le faire à 6h00 du matin. Et on observe à partir de 18h00 la publication en ligne par le journal *Ouest France* des nombreux articles de ses multiples éditions<sup>1</sup>. Cette automatisation de la publication en ligne concerne majoritairement les articles non classés dans des événements, et en particulier ceux publiés par les journaux locaux.

Assez logiquement, on ne retrouve pas une telle stratégie de publication pour les médias dont le support hors ligne n'impose pas les mêmes contraintes qu'à la presse écrite. Autre spécificité : l'AFP, qui publie relativement plus de documents pendant la nuit que les autres médias, tirant avantage de la présence de ses journalistes aux quatre coins du globe.

## En France en 2013 : 25 000 événements

L'événement médiatique est un concept central pour l'observation de la propagation de l'information. Abordé par les sociologues, les philosophes, les économistes et tout un chacun sous un angle différent, ce concept admet une variété de définitions (**ENCADRÉ**).

<sup>1</sup> Dans notre corpus, les articles de *Ouest France* ont été captés *a posteriori* en utilisant les pages d'archives du site internet. Il ne faisait pas partie du périmètre initial du projet OTMedia. Nous avons donc l'intégralité du contenu de 2013 et non uniquement ce qui était exposé à l'époque par les flux RSS comme nous l'avons fait pour les autres titres de Presse Quotidienne Régionale.

## Qu'est-ce qu'un événement ?

Dans le cadre de ce travail de recherche, nous avons généré la liste de l'ensemble des « événements » qui se sont déroulés en France en 2013 en utilisant un algorithme de détections d'événements basé sur la proximité linguistique. Mais qu'est-ce qu'un événement ?

Dans son approche la plus basique, un événement (au sens journalistique) relate un ou des faits qui sont considérés comme assez importants pour un public visé. En d'autres termes, un événement, c'est ce qui arrive, ce qui se produit et vaut la peine d'être rapporté à un public distant.

Il existe cependant de nombreuses autres approches de la notion d'événement. Nous en proposons ici quelques-unes.

### **L'événement comme miroir déformant d'une singularité de la réalité**

– L'approche phénoménologique tend à envisager l'événement comme une perturbation du monde. Michel Foucault (*Leçons sur la volonté de savoir*, Paris, « Hautes Études », Seuil/Gallimard, 2011) met en avant l'ubiquité du concept dans la définition qu'il en donne : « Un événement, c'est toujours une dispersion ; une multiplicité. C'est ce qui passe ici et là, il est polycéphale. » À la suite de Mai 68, Pierre Nora (« Le retour de l'événement » dans P. Nora et J. Le Goff (dir.), *Faire de l'histoire, tome 1*, Paris, Gallimard, 1974) ajoute à cette approche la dimension du récit porté par les mass media : « L'information colle à l'événement au point d'en faire partie intégrante. Non qu'elle le crée artificiellement, mais elle le constitue. »

### **L'événement comme objet sociétal acteur de changement**

– D'après Cyril Lemieux, Laurent Mucchielli et Érik Neveu (« Le sociologue dans le champ médiatique : diffuser et déformer ? », *Sociologie*, 1(2), 2010), l'événement se définit par des séquences d'action qui, marquées par une forte imprévisibilité, se traduisent en irréversibilités, c'est-à-dire déterminent de nouvelles orientations, sur des séquences d'actions ultérieures (voir également Michel Grossetti,

*Sociologie de l'imprévisible. Dynamiques de l'activité et des formes sociales*, Paris, PUF, 2004).

Selon Harvey Moloch et Marilyn Lester (« Informer : une conduite délibérée. De l'usage stratégique des événements », *Sociologie de la communication*, 1 : 433-451, 1997), trois acteurs constituent les événements : les promoteurs, les assembleurs et les consommateurs d'information. Ces événements peuvent être classés en événements de routine, en accidents, en scandales et en « heureux hasard ».

**L'événement médiatique** – L'AFP, de par ses pratiques, a élaboré une typologie des rubriques et des spécialisations thématiques pour le traitement des événements médiatiques (Éric Lagneau, « Le style agencier et ses déclinaisons thématiques. L'exemple des journalistes de l'Agence France Presse », *Réseaux*, 111 : 58-100, 2002). D'une certaine façon, cette conception de l'événement comme événement médiatique est celle qui se rapproche le plus du travail que nous avons mené ici.

### *Un algorithme de détection d'événements*

L'approche que nous avons choisie, du fait du nombre extrêmement important de documents à traiter, est une approche quantitative, basée sur la construction d'un algorithme de détection d'événements.

Dans sa version la plus simple, l'algorithme de détection d'événements consiste, de manière automatisée, à mettre ensemble tous les articles qui se rapportent au même « fait » et à exclure tous ceux qui ne s'y rapportent pas. De manière un peu plus technique, l'algorithme que nous avons développé utilise le titre, le chapeau, le corps du texte ainsi que la date de chaque article pour procéder à une analyse de vocabulaire lui permettant de regrouper ensemble les documents partageant un ensemble de mots fortement similaires (**ENCADRÉ**).

## L'algorithme de détection d'événements

Dans la communauté scientifique, les algorithmes de détection d'événements sont regroupés sous l'appellation anglaise de Topic Detection and Tracking (TDT). Ils font partie de la grande famille des algorithmes qui analysent des textes et essaient par diverses approches d'en extraire du sens. On parle de Traitement Automatique des Langues (TAL) ou Natural Language Processing. De très nombreuses approches ont été utilisées depuis quelques décennies pour tenter de résoudre ce problème<sup>1</sup>.

Notre algorithme de détection d'événements – dit « en ligne » – organise un flux constant d'articles, au fur et à mesure de leur apparition, dans des regroupements (clusters) cohérents. La consistance de ces regroupements est garantie aussi bien sémantiquement que temporellement afin de recouper la notion d'événement médiatique. Les principales étapes sont les suivantes :

1. Nous appliquons sur le texte quelques pré-traitements qui ont pour objectif de rendre les résultats robustes. Nous supprimons tout d'abord des documents les mots très communs de la langue française qui n'apportent aucune information particulière (environ 250 mots dont par exemple : alors, cela, mais, ici, etc.). Nous ne conservons ensuite que la racine des mots en appliquant un algorithme de « racinisation » ou stemming. Ainsi les comparaisons entre mots de la même famille sont facilitées.

2. Chaque article est dans un deuxième temps décrit par un vecteur, au sens mathématique, en se basant sur son titre et son contenu. Ce vecteur sémantique mesure l'importance relative de chacun des mots de l'article par rapport à l'ensemble du corpus. Une approche classique pour former ce vecteur est d'utiliser la représentation TF-IDF (Term Frequency – Inverse Document Frequency). Ainsi,

---

<sup>1</sup> On pourra en retrouver une bonne synthèse dans l'article d'Allan J., Harding S., Fisher D., Boliver A., Guzman-Lara S. et Amstutz P. (2005). « Taking Topic Detection From Evaluation To Practice », Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05).

l'importance d'un mot dépend de sa fréquence d'apparition dans le document, mais est pondérée inversement par le nombre de documents dans lequel il apparaît. De plus nous accordons un bonus (facteur multiplicatif de 5) aux mots apparaissant dans le titre puisqu'ils sont supposés décrire précisément le contenu de l'article.

3. À ce stade, chaque article est donc représenté par un vecteur mathématique. Nous les agglomérons successivement en fonction de leur similarité sémantique. La similarité entre deux articles est obtenue en mesurant la distance qui sépare leurs deux vecteurs sémantiques. Là encore, plusieurs approches et de nombreuses fonctions existent pour mesurer les distances entre vecteurs. Les vecteurs sémantiques ayant ici plusieurs centaines de milliers de dimensions (une pour chaque mot), la méthode la plus adaptée consiste à mesurer l'angle entre les deux vecteurs. Il s'agit de la distance dite « cosinus » que nous utilisons ici.

4. L'algorithme d'agglomération itérative regroupe ainsi à chaque étape les deux articles les plus proches au sein d'un événement. Le processus est stoppé lorsque la distance entre deux articles devient trop grande et dépasse un certain seuil. Ce seuil est déterminé empiriquement en se basant sur des événements créés manuellement.

5. Nous considérons qu'un événement est finalisé s'il ne reçoit plus de nouvel article pendant une certaine période de temps. Dans notre cas, nous utilisons une fenêtre glissante de 24 heures. Ainsi, un événement qui serait couvert le lundi et le mardi puis le samedi (sans aucun article du mercredi au vendredi) produira deux regroupements distincts.

6. Finalement, pour qu'il y « ait » événement, nous avons imposé deux seuils sur les regroupements issus de l'algorithme de détection : d'une part, qu'au moins deux médias couvrent le « fait » ; d'autre part, que le « fait » soit relaté dans un minimum de dix articles sur l'une des journées. Bien sûr, ces seuils peuvent être sujets à discussion. Ils n'affectent cependant pas quantitativement nos résultats.

Dans quelle mesure notre algorithme nous permet-il véritablement de mettre à jour les événements qui se sont déroulés en 2013 ? Regrouper de manière automatique des documents dans des événements nous confronte à deux types d'erreur : d'une part, les « faux positifs », qui sont des articles inclus dans un événement mais qui ne devraient pas l'être car ils ne traitent pas en fait de cet événement ; d'autre part, les « faux négatifs », c'est-à-dire les articles qui n'ont par erreur pas été inclus à l'intérieur d'un événement dont ils traitent pourtant. Toute la difficulté repose dans la détermination des seuils adéquats de proximité linguistique afin de limiter tout à la fois le nombre de « faux positifs » et de « faux négatifs ».

Notre algorithme de détection d'événements fait aussi bien dans la détermination d'événements et la classification de documents à l'intérieur de ces événements que les algorithmes considérés aujourd'hui comme l'état de l'art<sup>1</sup>. De plus, nous avons comparé notre liste d'événements à celle produite par l'explorateur d'actualités européen EMM (European Media Monitoring). Cet « explorateur » publie chaque jour les 19 événements ayant fait l'objet du plus grand nombre d'articles. Plus de 92% des événements identifiés comme tels par EMM en 2013 apparaissent également dans la liste d'événements que nous avons identifiés.

### *Les événements : quelques éléments de description*

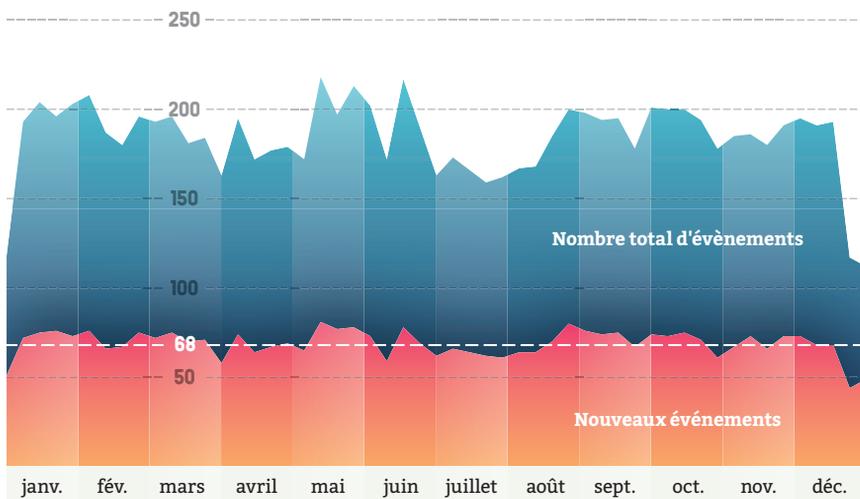
L'application de notre algorithme de détection d'événements au corpus décrit plus haut nous a permis de mettre à jour, pour l'année 2013, 24 985 événements, soit en moyenne 68 *nouveaux* événements par jour (**FIGURE 9**).

Il est frappant de noter à quel point le nombre moyen de nouveaux événements est stable dans le temps, comme si chaque jour le même nombre d'événements se produisait, inlassablement, à l'exception des week-ends (et des vacances) où ceux-ci sont moins nombreux. On aurait pu, au contraire, s'attendre à ce que l'actualité soit variable dans son offre d'événements, certains jours étant plus intenses en

<sup>1</sup> Pour le vérifier, nous avons appliqué notre algorithme à un corpus échantillon fourni à cette fin par la communauté internationale en sciences informatiques : le « Topic and Detection Tracking Pilot Study Corpus ». Ce corpus comprend des données issues de médias sur lesquelles une « vérité terrain » a été effectuée : des humains ont identifié les événements de ce corpus. Ce corpus permet ainsi de comparer les résultats obtenus par des algorithmes à ceux produits par une analyse humaine de référence.

→ Figure 9

## 68 NOUVEAUX ÉVÈNEMENTS PAR JOUR EN MOYENNE



› Sur cette figure, les données ont été extrapolées du 1<sup>er</sup> au 31 juillet afin de ne pas faire apparaître visuellement les conséquences de la panne informatique décrite dans le texte et qui a affecté la constitution de notre corpus au cours de l'été 2013. Bien sûr, cela n'est pas le cas dans le cadre de notre analyse et des statistiques présentées dans le corps du livre.

événements que d'autres, par le hasard des faits ; d'autres jours faisant la « grève des événements », pour détourner le mot heureux de Jean Baudrillard<sup>1</sup>. On aurait pu s'attendre à ce que, pour reprendre l'hypothétique annonce imaginée par Gaye Tuchman, les médias nous fassent savoir certains jours que « du fait de circonstances hors de leur contrôle, rien de nouveau ne s'est produit aujourd'hui » et qu'ils ne vont donc rien publier<sup>2</sup>. Mais non, au rythme de 68 en moyenne, les événements viennent toujours en nombre égal frapper chaque jour à la porte des journalistes. Car l'événement est ici *médiatique*, construit par le journaliste qui ne peut rendre copie blanche<sup>3</sup>.

Un événement dure en moyenne un peu moins de deux jours (42 heures), et comprend 34 articles publiés par 15 médias différents<sup>4</sup>.

1 Baudrillard, J. (1992). *L'illusion de la fin ou la grève des événements*, Collection Idées, Éditions Galilée.

2 Tuchman, G. (1978). *Making news*, *op. cit.*

3 La stabilité de ce nombre est d'autant plus étonnante que si, à la télévision, à la radio ou encore sur le papier, les journalistes sont soumis chaque jour à des contraintes d'espace et de temps très précises, Internet aurait pu leur offrir la liberté du choix de la quantité.

4 Ce que nous appelons « durée d'un événement » correspond à la durée de la couverture mé-

De quoi traitent les événements ? Pour répondre à cette question, nous les avons classifiés par leur thématique, en utilisant le système de classification de l'AFP qui distingue 17 thématiques (les « top-level media topics ») : (1) Art et Culture ; (2) Crime, Loi et Justice ; (3) Désastre et Accident ; (4) Économie, Business et Finance ; (5) Éducation ; (6) Environnement ; (7) Santé ; (8) Intérêts Humains ; (9) Emploi ; (10) Style de vie et Loisirs ; (11) Politique ; (12) Religion et Croyance ; (13) Sciences et Technologie ; (14) Société ; (15) Sport ; (16) Conflits, Guerre et Paix ; et (17) Météo<sup>1</sup>. Un événement peut être associé par l'AFP à plusieurs thématiques. C'est le cas de 8429 événements dans notre base de données. Par exemple quand, le 1<sup>er</sup> janvier 2013, l'AFP (première à couvrir l'événement en France) publie une dépêche à 8h18 annonçant que « le Sénat approuve l'accord évitant le mur budgétaire » aux États-Unis, cet événement est classé tout à la fois dans « Économie, Business et Finance » et dans « Politique ».

La grande majorité des événements se rapportent à la politique (30%) ; à l'économie et la finance (28) ; et au crime, à la loi et à la justice (22%). Le sport arrive en 4<sup>e</sup> position, avec une présence dans 12% des événements (**FIGURE 10**).

La faible présence de thématiques comme « Art et culture » ou « Sciences et Technologies » ne signifie pas que les médias ne publient pas d'articles liés à ces thématiques. Mais ces articles ne font le plus souvent pas partie d'« événements » tels que nous les avons définis, leur couverture comportant moins de 10 articles sur la journée (ou parce qu'ils ne sont traités que par un seul média). Ces thématiques sont en général considérées comme des nouvelles « non urgentes », plus spécifiquement des *soft news*, et s'inscrivent donc sur le long terme, à l'exception des manifestations spécifiques telles que les Nuits Blanches, la Fête de la Musique ou le Festival de Cannes.

Dans le prochain chapitre, nous présentons en détails trois événements : les écoutes de la NSA divulguées en exclusivité par le journal

---

diatique de l'événement – l'intervalle de temps entre le premier et le dernier article couvrant l'événement –, pas à la durée de l'événement lui-même.

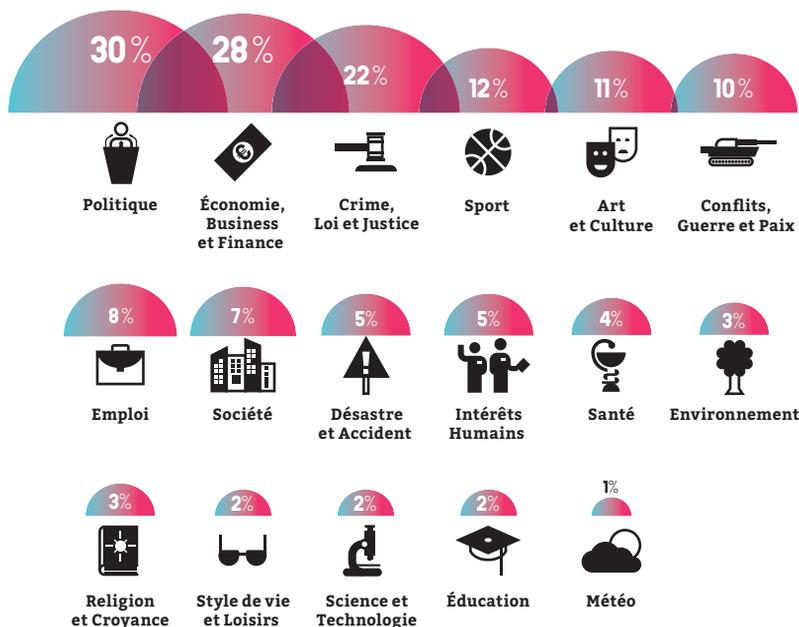
1 L'AFP utilise le standard IPTC, un système de classification international hiérarchique des sujets d'actualité, pour annoter l'intégralité de ses dépêches. Pour être tout à fait précis, pour définir la thématique de chaque événement, nous avons utilisé les métadonnées associées à la première dépêche AFP qui apparaît dans l'événement. Plus de 93% des événements dans notre corpus contiennent au moins une dépêche AFP. Nous n'avons pas déterminé la thématique des événements ne comprenant pas de dépêche AFP.

→ Figure 10

## UNE MAJORITÉ D'ÉVÉNEMENTS LIÉS À LA POLITIQUE

Répartition du nombre d'événements par thématiques

Plusieurs thématiques possibles par événement



*Le Monde*, l'enquête de la justice sur le micro-parti de Marine Le Pen révélée par le site *Mediapart*, et l'annonce par l'AFP de la participation de François Fillon à la primaire de droite (qui débouchera trois ans plus tard sur sa candidature à l'élection présidentielle).

### *La production d'information hors événements*

Sur les deux millions et demi d'articles présents dans notre base de données, seuls 33% sont classés à l'intérieur d'un événement. Ils représentent néanmoins un peu plus de 40% de la production totale d'information, ces articles étant en moyenne plus longs que les articles non classés. Les 67% restant relèvent de deux catégories.

D'une part, du fait du seuil minimal de deux médias que nous avons imposé pour la définition d'un événement, les actualités locales couvertes principalement par la presse départementale ou régionale ne donnent le plus souvent pas lieu à des événements (il est

d'ailleurs intéressant de noter que les médias locaux représentent 68% des documents non classés, alors qu'ils ne représentent « que » 57% des documents dans notre base de données.) Ainsi, un court article (1 421 caractères) publié par *Nice Matin* sur son site internet le 31 janvier 2013 sur la fermeture de la mairie du Plan-de-Grasse (« La mairie du Plan-de-Grasse fermée... pour vétusté ») peut être certes considéré comme couvrant un « événement » local, mais il ne sera classé dans aucun événement dans le cadre de notre analyse, la fermeture de cette petite mairie n'ayant donné lieu à aucune couverture par d'autres médias.

D'autre part, une partie importante du travail des journalistes consiste en l'écriture d'articles non rattachés directement à l'actualité, qu'il s'agisse de ce que l'on qualifie communément aujourd'hui de « *slow journalism* », d'articles de fond, de critiques de livres ou de réflexions sur le contexte sociétal. Michael Schudson parle à propos de ces articles d'« information contextuelle », Gaye Tuchman d'information « non urgente » par opposition à l'information que l'on pourrait qualifier de « chaude<sup>1</sup> ».

## La propagation de l'information vue à travers le prisme des événements

Qui produit l'information ? Qui la reprend et avec quelle réactivité ?

Pour répondre à ces questions, nous avons étudié les 25 000 événements qui ont eu lieu en 2013. Pour chacun des articles appartenant à ces événements, nous connaissons l'heure de publication. Nous sommes ainsi en mesure de déterminer le *news breaker* – le premier média à publier un article sur l'événement (**ENCADRÉ**) – puis d'établir une chronologie de l'événement médiatique. En particulier, nous pouvons étudier la réactivité des différents médias, ainsi que le nombre d'articles publiés par chaque média à l'intérieur de l'événement.

<sup>1</sup> Schudson, M. (1981) : *Discovering the News*, *op. cit.* Tuchman, G. (1978). *Making news*, *op. cit.*

### Qui sont les *news breakers* ?

L'approche du *breaking news* que nous avons adoptée est relativement large. Est défini comme *news breaker* d'un événement le premier média à avoir publié un article en ligne dans le cadre de cet événement.

Cette approche rejoint celle que l'on pourrait qualifier de plus « traditionnelle » du *breaking news* lorsque l'événement est le fruit d'un travail d'investigation et que le média *news breaker* est le premier à mettre à jour une information (par exemple les écoutes américaines de la NSA).

On peut interroger la validité d'une telle approche lorsque l'événement est attendu. En 2013, lors du conclave à l'issue duquel le pape François a été élu, des milliers de journalistes venus du monde entier attendaient ensemble de voir enfin apparaître la fumée blanche. Peut-on véritablement qualifier de *news breaker* le premier média à avoir publié sur son site internet le nom du cardinal apparu simultanément au balcon aux yeux des médias du monde entier ? De même, peut-on qualifier de *news breaker* le premier média à couvrir la conférence de presse d'un homme politique ou le communiqué de presse d'une entreprise ?

On peut bien entendu discuter cette terminologie. Mais il est intéressant de noter que si les utilisateurs attribuent une valeur positive au fait d'être informés en premier, alors de facto il n'est pas anodin pour un média d'avoir publié un article sur tel ou tel événement avant ou après ses concurrents.

Les médias vont vite. Ils copient beaucoup. Et ils ne créditent que très peu.

*Une vitesse de propagation extrêmement élevée*

La vitesse de propagation d'un événement en ligne est extrêmement rapide. Il faut en moyenne moins de trois heures (175 minutes) pour qu'un deuxième média couvre un événement déjà couvert par

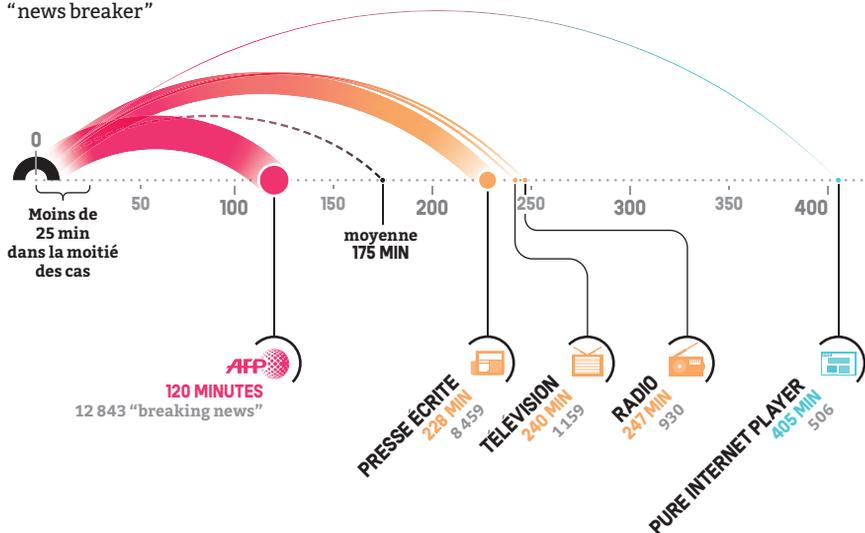
un premier média. Trois heures en moyenne, cela peut paraître long, mais cette statistique recouvre une forte disparité. En effet, un certain nombre d'événements se propagent beaucoup plus rapidement. La moitié des événements se propagent en moins de 25 minutes; un quart en seulement... 230 secondes, et 10% en seulement 4 secondes, ce dernier chiffre reflétant la publication automatique sur le site d'un certain nombre de médias de contenus déjà mis en forme par l'AFP.

La vitesse de propagation varie en fonction de plusieurs facteurs. Tout d'abord, elle varie en fonction du média *news breaker*. Nous venons de souligner la reprise automatique de contenus AFP; à l'autre extrême, le temps de réaction est le plus long quand le *news breaker* est un *pure internet player* (FIGURE 11). Il faut en moyenne 405 minutes (soit près de 7 heures) pour qu'une information révélée par un *pure player* soit reprise sur le site internet d'un média concurrent (ou par l'AFP). C'est près de deux fois et demie le temps moyen de réactivité. Comment expliquer cette « lenteur » relative? Elle provient en partie de l'arbitrage auquel sont confrontés les journalistes entre rapidité d'une part et véracité de l'autre. Car les *pure players*

→ Figure 11

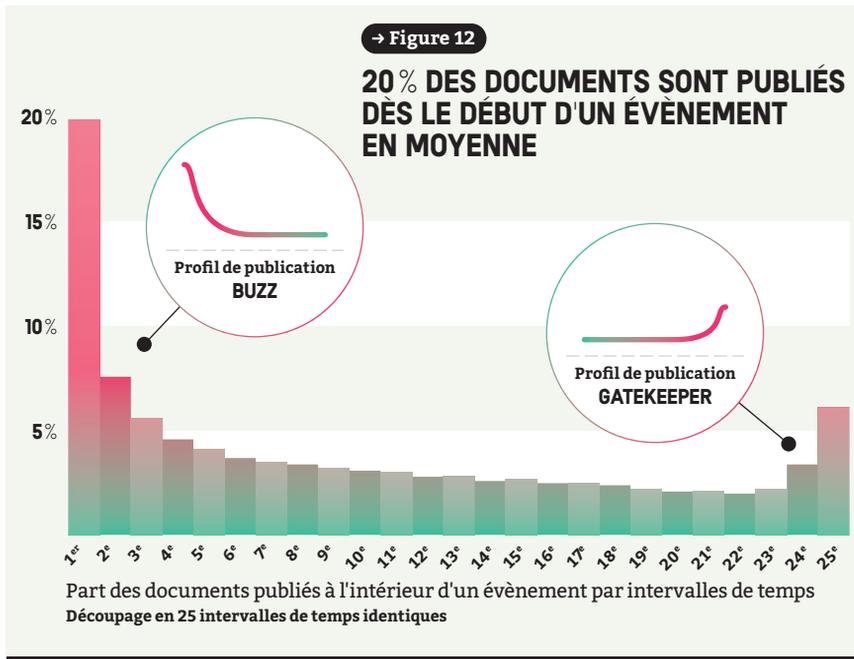
## MOINS DE 3H EN MOYENNE POUR QU'UNE INFORMATION SORTIE PAR UN MÉDIA SOIT REPRIS SUR LE SITE INTERNET D'UN CONCURRENT

Temps de réaction en minutes selon le support hors ligne du média "news breaker"



souffrent en 2013 comme aujourd'hui d'un déficit de réputation par rapport à un certain nombre de médias traditionnels; ceux-ci sont dès lors davantage incités à vérifier les informations publiées par les *pure players* plutôt qu'à les reprendre instantanément. Il suffit, par exemple, de penser au très faible nombre de reprises auquel a donné lieu l'interview de Ziad Takieddine dans *Mediapart* en novembre 2016 sur le financement libyen de la campagne de Nicolas Sarkozy.

Cela s'explique également par le type d'informations traitées par les *pure players*. Car il existe différents profils d'événements (le profil des événements correspond ici à ce que l'on appelle techniquement leur « distribution temporelle »). Pour les étudier, nous avons divisé chacun des événements dans notre base de données en 25 intervalles de longueur égale (ainsi pour un événement de 25 heures, la longueur de chaque intervalle est d'une heure; plus les événements durent longtemps, plus la longueur des intervalles est importante). La **FIGURE 12** représente en moyenne pour chaque événement le pourcentage des documents publiés à l'intérieur des différents intervalles.



Il est tout d'abord frappant de noter que la majorité des documents sont publiés au tout début de l'événement, près de 20% uniquement dans le premier intervalle. Cela traduit la forte réactivité des différents médias que nous avons soulignée plus haut, mais également le fait qu'en moyenne, dès qu'une information est publiée par un média, elle est reprise immédiatement – sous une forme proche du « buzz » – par l'ensemble des différentes publications. L'information semble ensuite perdre rapidement de son intérêt pour les médias.

Ce profil moyen de publication cache cependant une seconde réalité, qui apparaît dans les trois derniers intervalles où l'on observe une remontée de la part des documents publiés. En effet, au côté des événements à la forme de « buzz », on observe un autre profil d'événements que l'on pourrait qualifier, pour reprendre la terminologie initialement proposée par Kurt Lewin, d'événements « gatekeeper ». Autrement dit d'événements qui, pour être repris par la majorité des médias, ont d'abord besoin d'être couverts par un média de référence. Ainsi, quand, le 1<sup>er</sup> janvier 2013, *Le Dauphiné Libéré* publie à 06h05 sur son site internet un article sur un vol dans une bijouterie à la suite duquel le cambrioleur choisit de rendre le butin accompagné d'une boîte de chocolats (« Après un vol dans une bijouterie : il rend le butin... plus des chocolats ! »), seuls *Le Parisien* et *Direct Matin* reprennent l'information dans la journée. Jusqu'à la publication le lendemain d'un article sur le sujet sur le site internet de la radio RTL, donnant lieu ensuite à sept publications successives, par *Le Journal du Dimanche*, Europe1, le *Huffington Post*, *Sud Ouest*, BFM TV ou encore *Metro*.

Ces différents profils d'événements se reflètent également dans la durée des événements. Ainsi les événements « gatekeeper » sont en moyenne plus longs que les événements à la forme de « buzz » (l'événement du vol et de la boîte de chocolats a duré près de trois jours), notamment du fait de l'intervalle de temps entre la publication du premier article et celle de l'article par le média de référence qui transformera l'événement anecdotique en événement médiatique. Une fois que l'événement a franchi « la porte » du média de référence, l'ensemble des documents le couvrant sont publiés dans un intervalle de temps relativement court.

*La réactivité comme seul objectif ?*

Comment un journaliste (même le meilleur !) peut-il écrire un article de qualité sur un événement... en l'espace de seulement cinq minutes ? De deux choses l'une : soit le journaliste était déjà « sur l'événement » et il s'est fait griller la politesse par un autre journaliste travaillant pour un média concurrent<sup>1</sup>, soit ce journaliste, branché sur ses fils d'actualité et sur les sites internet des médias concurrents, s'est rapidement (et parfois librement !) inspiré de ce qui venait d'être publié ailleurs et a eu recours au copié-collé. Pablo Boczkowski qui a mené une véritable enquête ethnographique à l'intérieur de la rédaction des deux principaux quotidiens argentins<sup>2</sup>, a documenté avec beaucoup de précision cette part du travail des journalistes qui consiste, sur Internet, à surveiller de près les sites des concurrents pour être sûr de ne pas passer à côté d'une information qu'ils couvrent – et cette tendance à avoir parfois recours à la « facilité » (corollaire de la rapidité) et à reprendre presque tel quel l'article d'un concurrent.

Cette reprise s'accompagne parfois d'un relâchement de la vigilance quant à la correction formelle, sans parler des fameuses « boulettes » des « stagiaires bandeau » des chaînes d'information en continu.

La reprise de dépêches d'agence, et en particulier, en France, des dépêches de l'AFP, est plus habituelle, et fait même partie du contrat qui lie l'agence à ses clients. Si les médias s'abonnent au fil de l'AFP, c'est aussi pour pouvoir rapidement publier les dépêches de l'agence sur leur site. Ce qui donne le plus souvent lieu à la précision en bas de l'article publié « (Avec AFP) ».

1 Voir, récemment, la publication par un journaliste du *Washington Post*, David Fahrenthold, d'une vidéo... de NBC montrant Donald Trump tenir des propos outrageux envers les femmes. Seulement sept minutes après la publication de l'information – originale – sur le site du *Post*, elle a été reprise par NBC... qui s'est donc fait griller la politesse sur sa propre *breaking news*. Une question reste néanmoins ouverte ici : NBC avait-il jusque-là volontairement choisi de ne pas diffuser cette vidéo de l'une de ses « stars » ? D'après le *Washington Post* (« *A caller had a lewd tape of Donald Trump. Then the race to break the story was on* », Paul Fahri, 7 octobre 2016), NBC News s'apprêtait à diffuser la vidéo...

2 Voir Boczkowski, P. J. (2010). *News at Work : Imitation in an Age of Information Abundance*, Chicago, University of Chicago Press.

### **Les nouvelles erronées : le côté obscur de la réactivité ?**

Être « le premier sur l'information » peut être vu aujourd'hui, à l'heure de l'information en continu, comme l'alpha et l'omega d'un grand nombre de médias. Cette stratégie a ses raisons d'être, mais elle ne devrait pas se faire à n'importe quel prix.

Faire le choix de ne pas être le premier sur l'information – donc, en apparence du moins, d'une moindre réactivité –, c'est aussi faire le choix de vérifier deux fois une information avant de la publier et d'éviter ainsi la publication de nouvelles erronées comme cela arrive certes rarement, mais malheureusement encore trop souvent.

C'est aussi parfois également faire le choix d'éviter les manquements à la déontologie, comme cela a été parfois le cas dans l'actualité récente, dans le cadre de la couverture des multiples attentats.

Notre approche quantitative – et l'analyse automatisée de plus de 25 000 événements – a ses limites, dont celle de ne pas pouvoir distinguer entre « bonne » et « mauvaise » réactivité. Cela concerne un nombre extrêmement réduit d'événements, mais il nous semble toutefois utile de préciser que nous avons bien conscience qu'être le premier sur une « information » n'est pas toujours une bonne chose. Surtout lorsque cette information se trouve être erronée.

### *L'importance du copié-collé*

Au-delà de ce sentiment partagé de l'importance croissante de l'homogénéisation des contenus à l'heure d'Internet, il est nécessaire de quantifier très précisément l'importance du copié-collé dans la production d'information en ligne.

Pour ce faire, nous avons développé un algorithme de détection de copie. Cet algorithme nous permet, pour chaque article publié à l'intérieur d'un événement, de le comparer à l'ensemble des articles qui ont été publiés avant lui dans cet événement, et de distinguer

le contenu qui n'est que du copié-collé de contenu précédemment publié du contenu original (**ENCADRÉ**).

### L'algorithme de détection de copie

Les algorithmes qui détectent les portions de textes identiques entre deux documents existent depuis très longtemps en informatique. Ils sont un composant essentiel de nombreux autres traitements. Vous avez peut-être déjà utilisé cette fonctionnalité de votre logiciel de traitement de textes. Nous parlons bien ici de copié-collé et non pas de plagiat, notion plus générale qui implique une reformulation potentielle du texte d'origine.

Nous avons développé une implémentation nous permettant de trouver rapidement, pour un article donné, l'ensemble des articles d'un corpus avec lesquels il partage du contenu. Afin d'accélérer les traitements, nous utilisons une technique de hachage (*hashing*) qui permet de représenter une donnée à l'aide d'une empreinte particulière. C'est cette empreinte qui sert à identifier rapidement les contenus similaires. Dans notre cas, nous avons « haché » les suites de cinq mots consécutifs (représentation classique en TAL, dénommée *n-gram*, et donc *5-grams* dans notre implémentation). Nous pouvons ainsi repérer les contenus identiques à partir du moment où ils sont constitués au minimum de cinq mots. Pour tout couple de textes présentant du contenu identique, nous agglomérons toutes ces suites de cinq mots pour former les blocs de texte identique. Pour considérer que deux textes présentent de la copie, nous imposons un seuil minimal de 100 caractères sur la taille d'au moins l'un de ces blocs<sup>1</sup>.

-----  
<sup>1</sup> On peut bien évidemment discuter de la « validité » de ce seuil de 100 caractères, que l'on pourrait vouloir un peu plus long ou bien un peu plus court. Le raccourcir, c'est prendre le risque de qualifier improprement de copié-collé des tournures qui sont simplement usuelles dans l'écriture journalistique. Le rallonger au contraire, c'est prendre le risque de ne pas considérer comme du copié-collé du contenu qui est effectivement du copié-collé. Ce qui nous intéresse principalement ici ce n'est pas tant la quantité de contenu copié-collé que la quantité de contenu original produit par le média, et que modifier faiblement ce seuil ne modifierait pas quantitativement les résultats que nous obtenons.

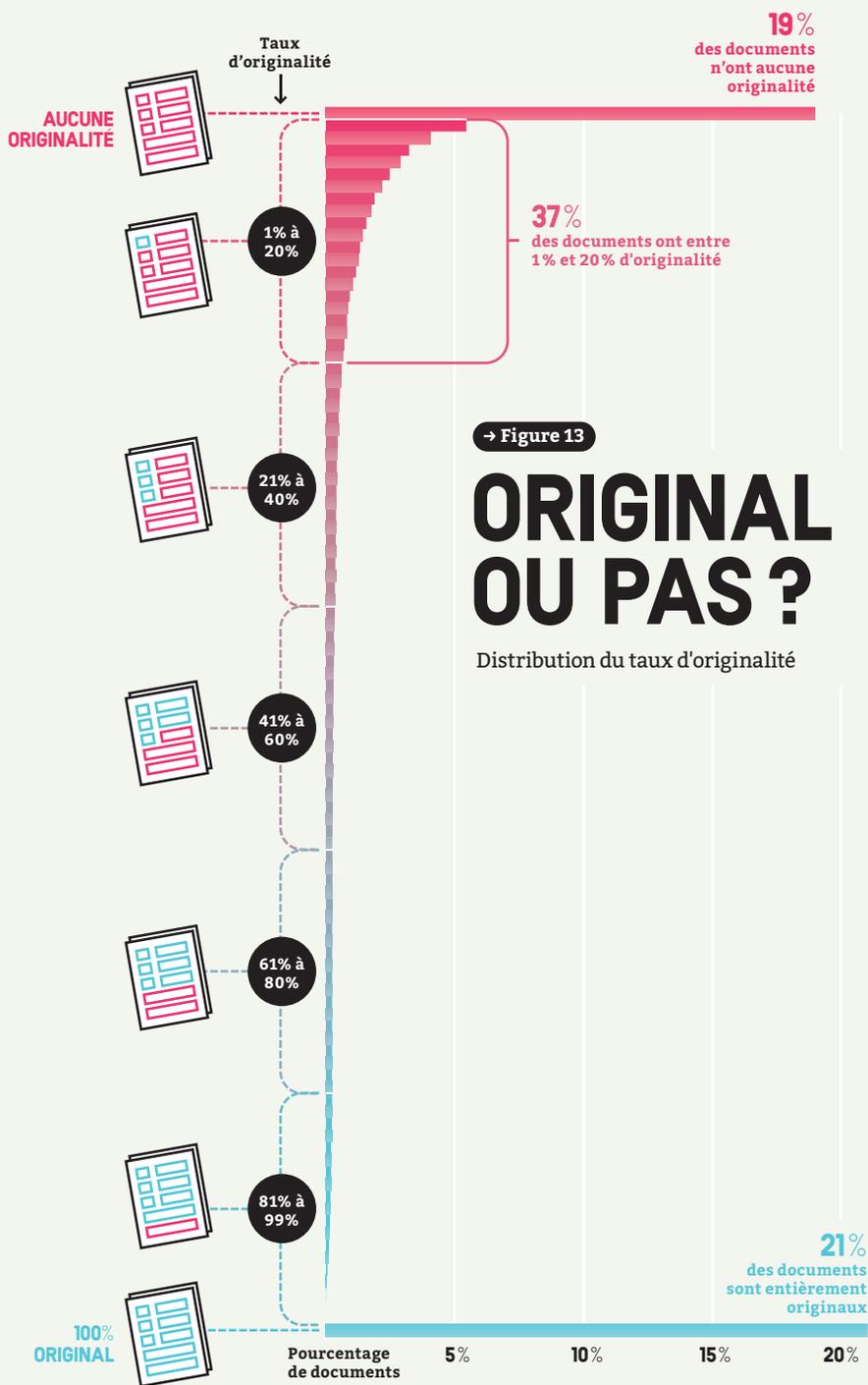
En appliquant cet algorithme au corpus formé par l'ensemble des articles d'un événement médiatique, nous pouvons mesurer plusieurs choses. En prenant les articles dans leur ordre chronologique, nous pouvons identifier pour chacun d'entre eux la proportion de contenu qui est originale et la part qui est copiée. Pour cette part copiée, nous distinguons ce qui est copié au sein du média (réutilisation de sa propre production) de ce qui est emprunté à un autre média. De plus, nous sommes également capables de dire dans quel article un bloc de texte est apparu pour la première fois et ainsi d'en attribuer la paternité au bon média.

Quelle est l'importance du contenu copié-collé par rapport au contenu original dans la production d'information sur Internet? D'après nos estimations, le taux moyen d'originalité d'un article en ligne – défini comme la part du contenu original de l'article rapporté à son contenu total – n'est que de 36%<sup>1</sup>. Autrement dit, 64% de ce qui est publié en ligne est du copié-collé pur et simple

De plus, ce taux moyen cache de très fortes disparités : la distribution du taux d'originalité est en fait bimodale, près de 19% des plus de 850 000 articles classés dans des événements ne présentant aucune originalité, et 21% d'entre eux étant entièrement originaux (**FIGURE 13**). Plus de la moitié des articles (56%) présentent moins de 20% de contenu original.

Le manque d'originalité s'explique toutefois en partie par de la copie interne. Lorsqu'ils couvrent un événement, les médias publient en effet en moyenne trois articles, deux si l'on exclut l'AFP dont la politique particulière de publication l'oblige à publier plusieurs dépêches par événement, d'abord une première dépêche courte, puis des dépêches plus longues. Au-delà du seul cas de l'AFP, on

<sup>1</sup> Rappelons qu'il s'agit d'une limite haute de l'estimation du taux d'originalité, puisque nous n'identifions dans ce travail que la copie et pas la reformulation. Cette approche « restrictive » de la notion de reprise de contenu sera étendue dans nos futurs travaux avec le développement d'algorithmes permettant d'évaluer avec précision la reformulation. De plus, nous n'identifions ici la copie qu'à l'intérieur des événements : si un événement médiatique est couvert à deux reprises mais à plusieurs jours d'intervalles et apparaît donc dans nos données comme deux événements distincts, alors nous estimons à tort comme original une partie du contenu des articles publiés dans le deuxième événement.



→ Figure 13

# ORIGINAL OU PAS ?

Distribution du taux d'originalité

observe une tendance chez les médias, suite à une première publication rapide (du fait de la recherche de la réactivité), à publier une seconde version plus complète des articles. À cela vient s'ajouter le recours croissant aux pratiques dites de « A/B testing » qui consistent à publier simultanément plusieurs versions d'un article (ou la même version d'un article avec différents titres) afin de sélectionner ensuite celui qui présente la meilleure « expérience utilisateur<sup>1</sup> ».

Mais parmi les 850 000 documents que nous avons classés dans des événements, plus de 620 000 (73%) présentent de la copie externe. Le taux moyen de copie externe est de 51,3% ; 72,6% conditionnellement au fait d'avoir recours à la copie. La copie est non seulement substantielle du point de vue du média qui copie, mais elle est également substantielle du point de vue du média copié. En moyenne, plus de 21% du contenu d'un média copié est copié. Ce taux dépasse les 82% si le média copié est le *news breaker*. Autrement dit, les médias reproduisent dans ce cas en intégralité plus des quatre cinquièmes des articles qu'ils reprennent ! Et ce chiffre ne provient pas uniquement de la copie des dépêches AFP. Si l'on considère les médias copiés autres que l'AFP, alors, conditionnellement au fait d'être copié, 60% du contenu du premier document dans l'événement est copié.

De plus, il est important de souligner à nouveau ici que notre analyse repose uniquement sur l'étude du contenu produit par les médias d'information. Ne sont pris en compte ni les documents publiés sur les réseaux sociaux ni ceux publiés par les agrégateurs (qui par définition « copient-collent » le titre, le chapeau et les premières phrases des articles auxquels ils renvoient).

Notons néanmoins que l'importance du copié-collé a été mesurée à l'intérieur des événements. Or moins de la moitié des articles publiés en ligne se rattachent de fait à des événements. Même si cela revient sans doute en partie à la surestimer<sup>2</sup>, on peut penser que l'originalité de ces articles de long court, qui ne s'inscrivent pas dans ces événements, est beaucoup plus proche de 100%.

1 Le *Washington Post* est sans doute le média d'information qui a le plus investi dans le « A/B testing » au cours des dernières années. Il a ainsi récemment lancé « Bandito », un système de testing de contenu en temps réel. Un tel système permet par exemple d'illustrer un même article avec deux photos différentes et de sélectionner ensuite de manière automatisée l'article dont la photo aura attiré le plus de clics.

2 Il suffit de penser par exemple aux fameux « marronniers » et autres classements d'hôpitaux ou d'universités. Mais il s'agit dans ce cas plus souvent de copie « interne » au média que de copie « externe » (d'autres médias).

Avant de tirer des conclusions sur ces chiffres de copie extrêmement élevés, il nous faut cependant mettre à part le cas de l'AFP. La reprise systématique du contenu de l'AFP est intéressante car elle s'inscrit dans une logique croissante d'homogénéisation de la production de contenus en ligne. Toutefois, elle n'est pas problématique du point de vue du droit de la propriété intellectuelle, les médias abonnés étant autorisés à reprendre s'ils le souhaitent en intégralité sur leur site internet le contenu des dépêches (et c'est d'ailleurs pour cette raison qu'ils font le choix de s'abonner à l'agence). La reprise systématique de contenus produits par des médias autres que l'AFP pose, elle, des questions en termes de respect de la propriété intellectuelle. Nous avons donc pour finir distingué la copie externe lorsque le média copié est l'AFP de la copie externe lorsque le média copié est un média autre que l'AFP. Les résultats sont frappants. Même une fois l'AFP exclue, le taux moyen de copie externe est de plus de 41% conditionnellement au fait d'avoir recours à la copie.

#### *Le risque croissant de l'homogénéisation*

Doit-on être surpris par de telles estimations ? Probablement pas. D'une part, ce résultat est cohérent avec les principaux résultats de la littérature existante, et en particulier avec les travaux de P. J. Boczkowski, qui a mis en évidence l'homogénéisation de la production d'actualités en ligne en Argentine à partir d'une approche qualitative<sup>1</sup>. D'autre part et surtout, le copié-collé est la contrepartie de la réactivité. On ne peut pas attendre d'un journaliste qu'il écrive à partir de rien plus de 2 000 caractères en l'espace de cinq minutes.

Doit-on s'en inquiéter ? L'enjeu est double ici. On observe une homogénéisation croissante des contenus ; il est facile de la dénoncer, mais l'on peut également décider d'observer le verre à moitié plein et noter que l'un des corollaires de ce recours à la copie est que davantage d'information est mise à disposition du plus grand nombre gratuitement aujourd'hui<sup>2</sup>.

1 Boczkowski, P. J., *News at Work*, *op. cit.*

2 De plus, l'homogénéisation croissante des contenus peut sembler être un moindre risque alors que de plus en plus nombreux sont ceux qui dénoncent les risques de polarisation politique liés à Internet. Cass Sunstein parle ainsi de « ségrégation idéologique » en ligne liée au fait que de plus en plus de citoyens sur Internet vivraient dans ce qu'il qualifie d'« *echo chambers* » ou de « *information cocoons* » (Sunstein, C. R. (2002). *Republic.com*, Princeton University Press; Sunstein, C. R. (2009). *Republic.com 2.0*, Princeton University Press); Robert Putnam de « cyberbalkanisation » (Putnam, R. D. (2000). *Bowling Alone : The Collapse and Revival of Ameri-*

Par ailleurs – et c'est ce qui nous semble ici le plus inquiétant –, la combinaison du recours à la copie et de la vitesse de propagation de l'information risque de tuer à terme les incitations des médias à produire une information originale, à faire perdre à l'information originale toute valeur commerciale. Or l'information originale est coûteuse à produire, et la copie mêlée à la réactivité fait de chaque site internet un substitut du site concurrent. Les médias n'arrivent ainsi plus à monétiser cette information si coûteuse à produire. C'est pourquoi il nous a semblé urgent de mettre à plat la réalité de la propagation de l'information en ligne, afin de souligner les difficultés qu'elle fait poser sur le futur de l'information et de nos démocraties.

---

*can Community*, Simon & Schuster, New York.). Notons néanmoins que cette vision d'Internet comme facteur de polarisation est loin d'être universellement partagée. Deux économistes américains, Matthew Gentzkow et Jesse Shapiro, ont montré que les relations sur Internet étaient moins ségréguées idéologiquement que les relations d'amitiés, de travail ou à l'intérieur de la famille (Gentzkow, M. et Shapiro, J. M. (2011). « Ideological Segregation Online and Offline », *The Quarterly Journal of Economics*, 126(4): pp. 1799-1839). Le débat est ouvert !



## CHAPITRE 2

### LA PROPAGATION DE L'INFORMATION SUR INTERNET : ILLUSTRATIONS

Trois événements ont été choisis pour leur représentativité parmi les 25 000 de notre corpus, afin d'illustrer un phénomène de propagation proche de la viralité : les écoutes de la NSA dévoilées par le journal *Le Monde* en octobre 2011, l'enquête de la justice sur le micro-parti de Marine Le Pen révélée par *Mediapart*, et enfin l'annonce par l'AFP de la candidature de François Fillon à la primaire de la droite.

#### ***Le Monde*, Snowden et les écoutes de la NSA**

Le 21 octobre 2013, le journal *Le Monde* révèle en exclusivité mondiale sur son site internet l'ampleur des écoutes de la NSA (l'agence nationale de sécurité américaine) dans le monde et en particulier en France. D'après les informations du quotidien, plus de 70 millions d'enregistrements de données téléphoniques françaises ont été effectués par l'agence entre le 10 décembre 2012 et le 8 janvier 2013. Ce scoop mondial, titré « Comment la NSA espionne la France » est d'abord publié sur le site du quotidien dans un article de 5 379 caractères à 6h01 passé de 13 secondes. Presque simultanément (10 secondes plus tard), *Le Monde* publie un deuxième article sur le sujet (de 5 768 caractères) intitulé « L'ampleur de l'espionnage mondial par la NSA ». Ces deux articles sont entièrement originaux, autrement dit, ils ne présentent aucune copie entre eux. *Le Monde* publie enfin sur son site internet un troisième article une demi-heure plus tard (à 6h32 passé de 52 de secondes) détaillant la nature des écoutes qui ont eu lieu (« Les services secrets américains très intéressés par Wanadoo et Alcatel-Lucent »).

Affaire d'État et d'espionnage couplée à l'histoire humaine d'un lanceur d'alerte<sup>1</sup>, Edward Snowden, les écoutes de la NSA sont un cas d'école de l'investigation journalistique à l'heure du numérique. Ce scoop repose en effet sur l'analyse systématique de documents confidentiels que *Le Monde* s'est procuré et dont certains éléments avaient été précédemment évoqués par l'hebdomadaire allemand *Der Spiegel* et le quotidien britannique *The Guardian*. Cas d'école de l'investigation, quelles leçons peut-on tirer de ce scoop quant à la propagation de l'information sur Internet ?

Tout juste 30 secondes après la publication du premier article sur lemonde.fr (à 6h01 passé de 43 secondes), l'AFP – qui était sous embargo – produit une première courte dépêche (494 caractères) titrée « La NSA a récolté des millions de données en France », et empruntée pour l'essentiel à l'article du quotidien (les parties copiées-collées de l'article du *Monde* sont reproduites en gras) :

« L'agence américaine de renseignement, NSA, a intercepté de façon massive **les communications téléphoniques des citoyens français**, révèle lundi le quotidien *Le Monde*, citant des documents de l'ancien consultant de la NSA, Edward Snowden. **Sur une période de trente jours**, entre le 10 décembre 2012 et le **8 janvier 2013, 70,3 millions d'enregistrements de données téléphoniques des Français ont été effectués par la NSA**, précise le site lemonde.fr avant la publication de la version papier. »

Cette dépêche de l'AFP, dont 40% du contenu est donc copié-collé de l'article du *Monde*, est classée dans les thématiques « Média » (incluse dans Art et Culture), « Informatique et technologie de l'information » (incluse dans Économie, Business et Finance) et Politique.

Cinq secondes plus tard, l'AFP publie une seconde dépêche, plus longue (3 177 caractères) dont près des trois quarts du contenu sont cette fois directement copiés-collés de l'article du *Monde* (la **FIGURE 14** illustre visuellement l'ampleur de la copie). Dans sa première comme dans sa seconde dépêche, l'AFP mentionne à de multiples reprises *Le Monde* comme la source l'information (« révèle lundi le quotidien Le Monde », « indique Le Monde », « d'après Le Monde »...).

<sup>1</sup> Sur les lanceurs d'alerte, on pourra lire l'excellent livre de Geoffroy de Lagasnerie : Lagasnerie, G. de (2015). *L'art de la révolte. Snowden, Assange, Manning*, Paris, Fayard.

Une demi-heure plus tard, le premier média (si l'on ne considère pas l'AFP) à couvrir en ligne les écoutes de la NSA est la radio RTL, qui publie sur son site internet à 6h29 un article 2 976 caractères (« Espionnage : la NSA a récolté des millions de données en France »). 81% du contenu de cet article est repris directement de la seconde dépêche de l'AFP. Quand, 12 minutes plus tard, *Le Nouvel Observateur* met à son tour en ligne un article de 3 526 caractères sur le sujet, la copie atteint cette fois les 89%. Rapidité et copié-collé caractérisent donc ainsi la diffusion en ligne de ce scoop mondial.

Au final, ce sont 119 articles (115 si l'on ne prend pas en compte ceux publiés par *Le Monde*) sur les écoutes de la NSA qui seront mis en ligne par 52 médias dans la seule journée du 21 octobre. Dans les trois heures qui suivent la publication du premier article par *Le Monde*, on trouve déjà sur Internet 53 papiers sur le sujet. Les Français dont le réveil sonne à 7 heures du matin peuvent indifféremment s'informer sur ces écoutes sur les sites du *Monde*, mais également de RTL, du *Nouvel Observateur* ou encore du *Figaro* ; ceux qui se lèvent à 8 heures n'ont que l'embaras du choix entre *Le Point*, France Info, *Le Parisien*, *Rue89*, *Les Échos*... pour n'en citer que quelques-uns. Certes les articles publiés sur tous ces sites ne présentent que très peu d'information originale, comme l'illustre la **FIGURE 15**. Mais au final, puisqu'ils reprennent en très grande partie le verbatim du *Monde*, le lecteur est indifférent quant au site sur lequel il doit se rendre.

Est-ce une bonne ou une mauvaise chose ? Cela reflète l'homogénéisation croissante des contenus produits sur Internet. Dans ce cas très précis, si *Le Monde* souffre potentiellement de la concurrence en ligne d'autres médias qui donnent à lire ses propres mots – au final 78% du document du premier article du *Monde* aura été repris par au moins un autre média ! –, il bénéficie en contrepartie d'un effet de réputation important, car ces autres médias le citent très massivement comme l'auteur du scoop.

La **FIGURE 16** illustre l'importance des citations reçues par *Le Monde* qui se trouve au centre de l'événement. Sur les 115 articles publiés par des médias concurrents, 87% y font référence comme étant la source de l'information. Il n'y est cependant pas seul, puisque de nombreux médias font le choix de citer également l'AFP. Un certain nombre de médias se réfèrent également à Europe 1 : Manuel Valls, alors ministre de l'Intérieur, se trouve être le 21 octobre l'invité de 8h15 de la matinale de la station. Interrogé par Jean-

## COMMENT LA NSA ESPIONNE LA FRANCE

Publication : 21/10/2013 06:01:13 - Taille : 5379 - Copie : 3,70 %

*Le Monde |21.10.2013 à 06h52| Par Jacques Follorou et Glenn Greenwald (Journaliste)*

L'avenir dira peut-être, un jour, pourquoi Paris est resté si discret, par rapport à Berlin ou Rio après les révélations sur les programmes d'espionnage électronique américain dans le monde. Car la France a été tout autant ciblée et dispose aujourd'hui de preuves tangibles que ses intérêts sont quotidiennement visés. Selon les documents de l'Agence nationale de sécurité (NSA) obtenus par Le Monde, les communications téléphoniques des citoyens français sont, en effet, interceptées de façon massive. Ces pièces, dévoilées en juin par l'ex-consultant de l'agence Edward Snowden, décrivent les techniques utilisées pour capter illégalement les secrets ou la simple vie privée des Français. Certains éléments ont été évoqués par l'hebdomadaire allemand Der Spiegel et le quotidien britannique The Guardian. D'autres sont inédits. Parmi les milliers de documents soustraits à la NSA par son ex-employé figure un graphique qui décrit l'ampleur des surveillances téléphoniques réalisées en France. On constate que sur une période de trente jours, du 10 décembre 2012 au 8 janvier 2013, 70,3 millions d'enregistrements de données téléphoniques des Français ont été effectués par la NSA.

LES TROIS PARTIES L'agence dispose de plusieurs modes de collecte. Quand certains numéros de téléphone sont utilisés dans l'Hexagone, ils activent un signal qui déclenche automatiquement l'enregistrement de certaines conversations. Cette surveillance récupère également les SMS et leur contenu en fonction de mots-clés. Enfin, de manière systématique, la NSA conserve l'historique des connexions de chaque cible. Cet espionnage apparaît au titre du programme "US-985D". L'explicitation exacte de ce sigle n'a pas été fournie, à ce jour, par les documents Snowden ni par d'anciens membres de la NSA. A titre de comparaison, les sigles utilisés par la NSA pour le même type d'interception visant l'Allemagne sont "US-987LA" et "US-987LB". Cette série de numéros correspondrait au cercle qualifié par les Etats-Unis de "troisième partie" auquel appartiennent la France, l'Allemagne mais aussi l'Autriche, la Pologne ou encore la Belgique. "La deuxième partie" concerne les pays anglo-saxons historiquement proches de Washington, le Royaume-Uni, le Canada, l'Australie et la Nouvelle-Zélande, connus sous le nom des "Five Eyes". "La première partie" concerne, elle, les seize services secrets américains. Les techniques utilisées pour ces interceptions apparaissent sous les codes "DRTBOX" et "WHITEBOX". Leurs caractéristiques ne sont pas connues. Mais on sait que grâce au premier code, 62,5 millions de données téléphoniques sont collectés en France du 10 décembre 2012 au 8 janvier 2013 et que le second permet d'enregistrer sur la même période 7,8 millions d'éléments. Les documents donnent suffisamment d'explications pour penser que les cibles de la NSA concernent aussi bien des personnes suspectées de liens avec des activités terroristes que des individus visés pour leur simple appartenance au monde des affaires, de la politique ou à l'administration française. Le graphique de la NSA montre une moyenne d'interceptions de 3 millions de données par jour avec des pointes à presque 7 millions les 24 décembre 2012 et 7 janvier 2013. Mais du 28 au 31 décembre, aucune interception ne semble avoir été opérée. Cet apparent arrêt d'activité pourrait s'expliquer, notamment, par le délai nécessaire à la reconduction, fin décembre 2012, par le Congrès américain de la section 702 de la loi encadrant l'espionnage électronique à l'étranger. De même, rien n'apparaît les 3, 5 et 6 janvier 2013 sans que l'on puisse, cette fois-ci, avancer de raison plausible. De nombreuses questions se posent encore, à commencer par l'identité précise des cibles et les justifications d'une collecte si massive de données sur un territoire étranger, souverain et allié. Sollicitées, les autorités américaines n'ont pas souhaité



## LA NSA A RÉCOLTÉ DES MILLIONS DE DONNÉES EN FRANCE (PRESSE)

Publication : 21/10/2013 06:01:43 - Taille : 494 - Copie : 40,49%

L'agence américaine de renseignement, NSA, a intercepté de façon massive **les communications téléphoniques des citoyens français**, révèle lundi le quotidien Le Monde, citant des documents de l'ancien consultant de la NSA, Edward Snowden. **Sur une période de trente jours**, entre le 10 décembre 2012 et **le 8 janvier 2013**, **70,3 millions d'enregistrements de données téléphoniques des Français** ont été effectués par la NSA, précise le site lemonde.fr avant la publication de la version papier.

# Le Monde

## COMMENT LA NSA ESPIONNE LA FRANCE

Publication : 21/10/2013 06:01:13 - Taille : 5 379 - Copie : 45,16%

Le Monde |21.10.2013 à 06h52| Par Jacques Follorou et Glenn Greenwald (Journaliste)

L'avenir dira peut-être, un jour, pourquoi Paris est resté si discret, par rapport à Berlin ou Rio après les révélations sur les programmes d'espionnage électronique américain dans le monde. Car la France a été tout autant ciblée et dispose aujourd'hui de preuves tangibles que ses intérêts sont quotidiennement visés. Selon les documents de l'Agence nationale de sécurité (NSA) obtenus par Le Monde, les communications téléphoniques des citoyens français sont, en effet, interceptées de façon massive. Ces pièces, dévoilées en juin par l'ex-consultant de l'agence Edward Snowden, décrivent les techniques utilisées pour capter illégalement les secrets ou la simple vie privée des Français. Certains éléments ont été évoqués par l'hebdomadaire allemand Der Spiegel et le quotidien britannique The Guardian. D'autres sont inédits. Parmi les milliers de documents soustraits à la NSA par son ex-employé figure un graphique qui décrit l'ampleur des surveillances téléphoniques réalisées en France. On constate que sur une période de trente jours, du 10 décembre 2012 au 8 janvier 2013, 70,3 millions d'enregistrements de données téléphoniques des Français ont été effectués par la NSA. LES TROIS PARTIES L'agence dispose de plusieurs modes de collecte. Quand certains numéros de téléphone sont utilisés dans l'Hexagone, ils activent un signal qui déclenche automatiquement l'enregistrement de certaines conversations. Cette surveillance récupère également les SMS et leur contenu en fonction de mots-clés. Enfin, de manière systématique, la NSA conserve l'historique des connexions de chaque cible. Cet espionnage apparaît au titre du programme "US-985D". L'explicitation exacte de ce sigle n'a pas été fournie, à ce jour, par les documents Snowden ni par d'anciens membres de la NSA. A titre de comparaison, les sigles utilisés par la NSA pour le même type d'interception visant l'Allemagne sont "US-987LA" et "US-987LB". Cette série de numéros correspondrait au cercle qualifié par les Etats-Unis de "troisième partie" auquel appartiennent la France, l'Allemagne mais aussi l'Autriche, la Pologne ou encore la Belgique. "La deuxième partie" concerne les pays anglo-saxons historiquement proches de Washington, le Royaume-Uni, le Canada, l'Australie et la Nouvelle-Zélande, connus sous le nom des "Five Eyes". "La première partie" concerne, elle, les seize services secrets américains. Les techniques utilisées pour ces interceptions apparaissent sous les codes "DRTBOX" et "WHITEBOX". Leurs caractéristiques ne sont pas connues. Mais on sait que grâce au premier code, 62,5 millions de données téléphoniques sont collectés en France du 10 décembre 2012 au 8 janvier 2013 et que le second permet d'enregistrer sur la même période 7,8 millions d'éléments. Les documents donnent suffisamment d'explications pour penser que les cibles de la NSA concernent aussi bien des personnes suspectées de liens avec des activités terroristes que des individus visés pour leur simple appartenance au monde des affaires, de la politique ou à l'administration française. Le graphique de la NSA montre une moyenne d'interceptions de 3 millions de données par jour avec des pointes à presque 7 millions les 24 décembre 2012 et 7 janvier 2013. Mais du 28 au 31 décembre, aucune interception ne semble avoir été opérée. Cet apparent arrêt d'activité pourrait s'expliquer, notamment, par le délai nécessaire à la reconduction, fin décembre 2012, par le Congrès américain de la section 702 de la loi encadrant l'espionnage électronique à l'étranger. De même, rien n'apparaît les 3, 5 et 6 janvier 2013 sans que l'on puisse, cette fois-ci, avancer de raison plausible. De nombreuses questions se posent encore, à commencer par l'identité précise des cibles et les justifications d'une collecte si massive de données sur un territoire étranger, souverain et allié. Sollicitées, les autorités américaines n'ont pas souhaité



## LA NSA A RÉCOLTÉ DES MILLIONS DE DONNÉES EN FRANCE (PRESSE)

Publication : 21/10/2013 06:01:48 - Taille : 3 177 - Copie : 74,69 %

L'agence américaine de renseignement, NSA, a intercepté de façon massive les communications téléphoniques des citoyens français, révèle lundi le quotidien Le Monde, citant des documents de l'ancien consultant de la NSA, Edward Snowden. Sur une période de trente jours, entre le 10 décembre 2012 et le 8 janvier 2013, 70,3 millions d'enregistrements de données téléphoniques des Français ont été effectués par la NSA, précise le site lemonde.fr avant la publication de la version papier. Ces pièces, dévoilées en juin par l'ex-consultant de l'agence, Edward Snowden, décrivent les techniques utilisées pour capter illégalement les secrets ou la simple vie privée des Français, ajoute le quotidien. La NSA dispose de plusieurs modes de collecte, indique Le Monde. Quand certains numéros de téléphone sont utilisés dans l'Hexagone, ils activent un signal qui déclenche automatiquement l'enregistrement de certaines conversations. Cette surveillance récupère également les SMS et leur contenu en fonction de mots-clés. Enfin, de manière systématique, la NSA conserve l'historique des connexions de chaque cible, précise le journal du soir. Cet espionnage apparaît au titre du programme "US-985D". D'après Le Monde, l'explicitation exacte de ce sigle n'a pas été fournie, à ce jour, par les documents Snowden ni par d'anciens membres de la NSA. A titre de comparaison, les sigles utilisés par la NSA pour le même type d'interception visant l'Allemagne sont "US-987LA" et "US-987LB", constate le quotidien. Cette série de numéros correspondrait au cercle qualifié par les Etats-Unis de "troisième partie" auquel appartiennent la France, l'Allemagne mais aussi l'Autriche, la Pologne ou encore la Belgique. "La deuxième partie" concerne les pays anglo-saxons historiquement proches de Washington, le Royaume-Uni, le Canada, l'Australie et la Nouvelle-Zélande, connus sous le nom des "Five Eyes". "La première partie" est constituée des seize services secrets américains. Les techniques utilisées pour ces interceptions apparaissent sous les codes "DRTBOX" et "WHITEBOX". Leurs caractéristiques ne sont pas connues, selon Le Monde, mais on sait que grâce au premier code, 62,5 millions de données téléphoniques sont collectés en France du 10 décembre 2012 au 8 janvier 2013 et que le second permet d'enregistrer sur la même période 7,8 millions d'éléments. Les documents donnent suffisamment d'explications pour penser que les cibles de la NSA concernent aussi bien des personnes suspectées de liens avec des activités terroristes que des individus visés pour leur simple appartenance au monde des affaires, de la politique ou à l'administration française. Le graphique de la NSA montre une moyenne d'interceptions de 3 millions de données par jour avec des pointes à presque 7 millions les 24 décembre 2012 et 7 janvier 2013, précise le quotidien. Par ailleurs, selon les documents de la NSA, obtenus par Le Monde, l'agence américaine s'est intéressée de près, entre le 1er et le 31 janvier 2013, aux adresses de messagerie wanadoo.fr, ancienne filiale d'Orange qui compte encore 4,5 millions d'utilisateurs, et alcatel.lucent.com, l'entreprise franco-américaine de télécommunications.



## LA NSA A RÉCOLTÉ DES MILLIONS DE DONNÉES EN FRANCE (PRESSE)

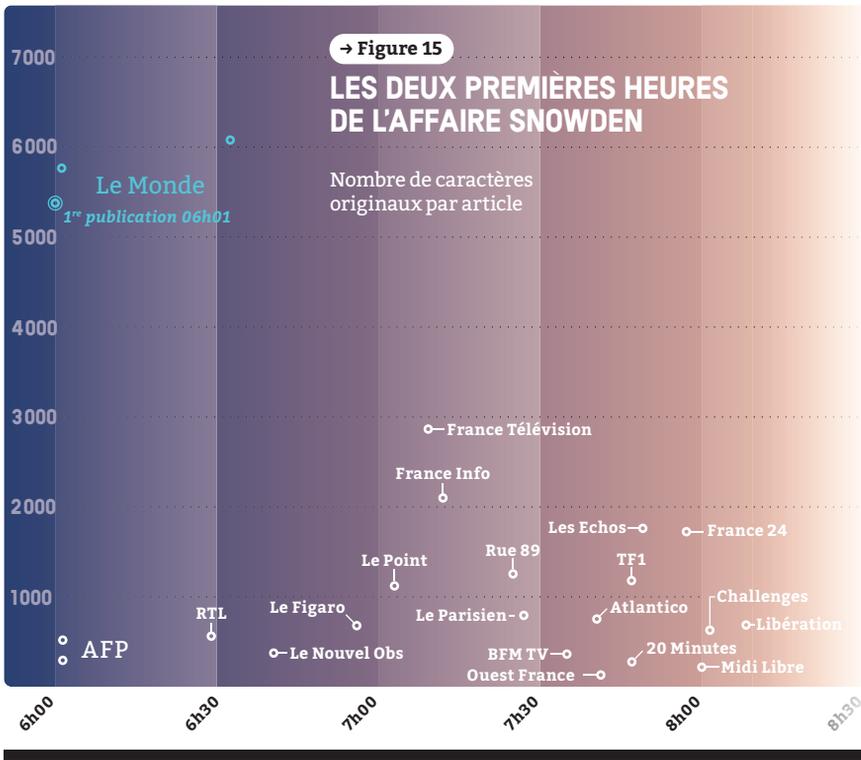
Publication : 21/10/2013 06:01:48 - Taille : 3 177 - Copie : 98,68 %

L'agence américaine de renseignement, NSA, a intercepté de façon massive les communications téléphoniques des citoyens français, révèle lundi le quotidien *Le Monde*, citant des documents de l'ancien consultant de la NSA, Edward Snowden. Sur une période de trente jours, entre le 10 décembre 2012 et le 8 janvier 2013, 70,3 millions d'enregistrements de données téléphoniques des Français ont été effectués par la NSA, précise le site [lemonde.fr](http://lemonde.fr) avant la publication de la version papier. Ces pièces, dévoilées en juin par l'ex-consulat de l'agence, Edward Snowden, décrivent les techniques utilisées pour capter illégalement les secrets ou la simple vie privée des Français, ajoute le quotidien. La NSA dispose de plusieurs modes de collecte, indique *Le Monde*. Quand certains numéros de téléphone sont utilisés dans l'Hexagone, ils activent un signal qui déclenche automatiquement l'enregistrement de certaines conversations. Cette surveillance récupère également les SMS et leur contenu en fonction de mots-clés. Enfin, de manière systématique, la NSA conserve l'historique des connexions de chaque cible, précise le journal du soir. Cet espionnage apparaît au titre du programme "US-985D". D'après *Le Monde*, l'explicitation exacte de ce sigle n'a pas été fournie, à ce jour, par les documents Snowden ni par d'anciens membres de la NSA. A titre de comparaison, les sigles utilisés par la NSA pour le même type d'interception visant l'Allemagne sont "US-987LA" et "US-987LB", constate le quotidien. Cette série de numéros correspondrait au cercle qualifié par les Etats-Unis de "troisième partie" auquel appartiennent la France, l'Allemagne mais aussi l'Autriche, la Pologne ou encore la Belgique. "La deuxième partie" concerne les pays anglo-saxons historiquement proches de Washington, le Royaume-Uni, le Canada, l'Australie et la Nouvelle-Zélande, connus sous le nom des "Five Eyes". "La première partie" est constituée des seize services secrets américains. Les techniques utilisées pour ces interceptions apparaissent sous les codes "DRTBOX" et "WHITEBOX". Leurs caractéristiques ne sont pas connues, selon *Le Monde*, mais on sait que grâce au premier code, 62,5 millions de données téléphoniques sont collectés en France du 10 décembre 2012 au 8 janvier 2013 et que le second permet d'enregistrer sur la même période 7,8 millions d'éléments. Les documents donnent suffisamment d'explications pour penser que les cibles de la NSA concernent aussi bien des personnes suspectées de liens avec des activités terroristes que des individus visés pour leur simple appartenance au monde des affaires, de la politique ou à l'administration française. Le graphique de la NSA montre une moyenne d'interceptions de 3 millions de données par jour avec des pointes à presque 7 millions les 24 décembre 2012 et 7 janvier 2013, précise le quotidien. Par ailleurs, selon les documents de la NSA, obtenus par *Le Monde*, l'agence américaine s'est intéressée de près, entre le 1<sup>er</sup> et le 31 janvier 2013, aux adresses de messagerie [wanadoo.fr](http://wanadoo.fr), ancienne filiale d'Orange qui compte encore 4,5 millions d'utilisateurs, et [alcatel.lucent.com](http://alcatel.lucent.com), l'entreprise franco-américaine de télécommunications.

## LES FRANÇAIS DANS LE VISEUR DE LA NSA

Publication : 21/10/2013 06:40:58 - Taille : 3 526 - Copie : 89,36 %

Cybersurveillance : la NSA partagerait ses données avec Israël. L'agence américaine de renseignement, NSA, a intercepté de façon massive les communications téléphoniques des citoyens français, révèle lundi 21 octobre le quotidien "Le Monde", citant des documents de l'ancien consultant de la NSA, Edward Snowden. Sur une période de trente jours, entre le 10 décembre 2012 et le 8 janvier 2013, 70,3 millions d'enregistrements de données téléphoniques des Français ont été effectués par la NSA, précise le site du quotidien avant la publication de la version papier. Ces pièces, dévoilées en juin par l'ex-consulat de l'agence, Edward Snowden, décrivent les techniques utilisées pour capter illégalement les secrets ou la simple vie privée des Français, ajoute le quotidien. La NSA dispose de plusieurs modes de collecte, indique "Le Monde". Quand certains numéros de téléphone sont utilisés dans l'Hexagone, ils activent un signal qui déclenche automatiquement l'enregistrement de certaines conversations. Cette surveillance récupère également les SMS et leur contenu en fonction de mots-clés. Enfin, de manière systématique, la NSA conserve l'historique des connexions de chaque cible, précise le journal du soir. La France dans la "troisième partie" cet espionnage apparaît au titre du programme "US-985D". D'après "Le Monde", l'explicitation exacte de ce sigle n'a pas été fournie, à ce jour, par les documents Snowden ni par d'anciens membres de la NSA. A titre de comparaison, les sigles utilisés par la NSA pour le même type d'interception visant l'Allemagne sont "US-987LA" et "US-987LB", constate le quotidien. Cette série de numéros correspondrait au cercle qualifié par les Etats-Unis de "troisième partie" auquel appartient la France, l'Allemagne mais aussi l'Autriche, la Pologne ou encore la Belgique. "La deuxième partie" concerne les pays anglo-saxons historiquement proches de Washington, le Royaume-Uni, le Canada, l'Australie et la Nouvelle-Zélande, connus sous le nom des "Five Eyes". "La première partie" est constituée des seize services secrets américains. Les techniques utilisées pour ces interceptions apparaissent sous les codes "DRTBOX" et "WHITEBOX". Leurs caractéristiques ne sont pas connues, selon "Le Monde", mais on sait que grâce au premier code, 62,5 millions de données téléphoniques sont collectés en France du 10 décembre 2012 au 8 janvier 2013 et que le second permet d'enregistrer sur la même période 7,8 millions d'éléments. 3 à 7 millions de données interceptées par jour, les documents donnent suffisamment d'explications pour penser que les cibles de la NSA concernent aussi bien des personnes suspectées de liens avec des activités terroristes que des individus visés pour leur simple appartenance au monde des affaires, de la politique ou à l'administration française. Le graphique de la NSA montre une moyenne d'interceptions de 3 millions de données par jour avec des pointes à presque 7 millions les 24 décembre 2012 et 7 janvier 2013, précise le quotidien. Par ailleurs, selon les documents de la NSA, obtenus par "Le Monde", l'agence américaine s'est intéressée de près, entre le 1<sup>er</sup> et le 31 janvier 2013, aux adresses de messagerie wanadoo.fr, ancienne filiale d'Orange qui compte encore 4,5 millions d'utilisateurs, et alcatel.lucent.com, l'entreprise franco-américaine de télécommunications.

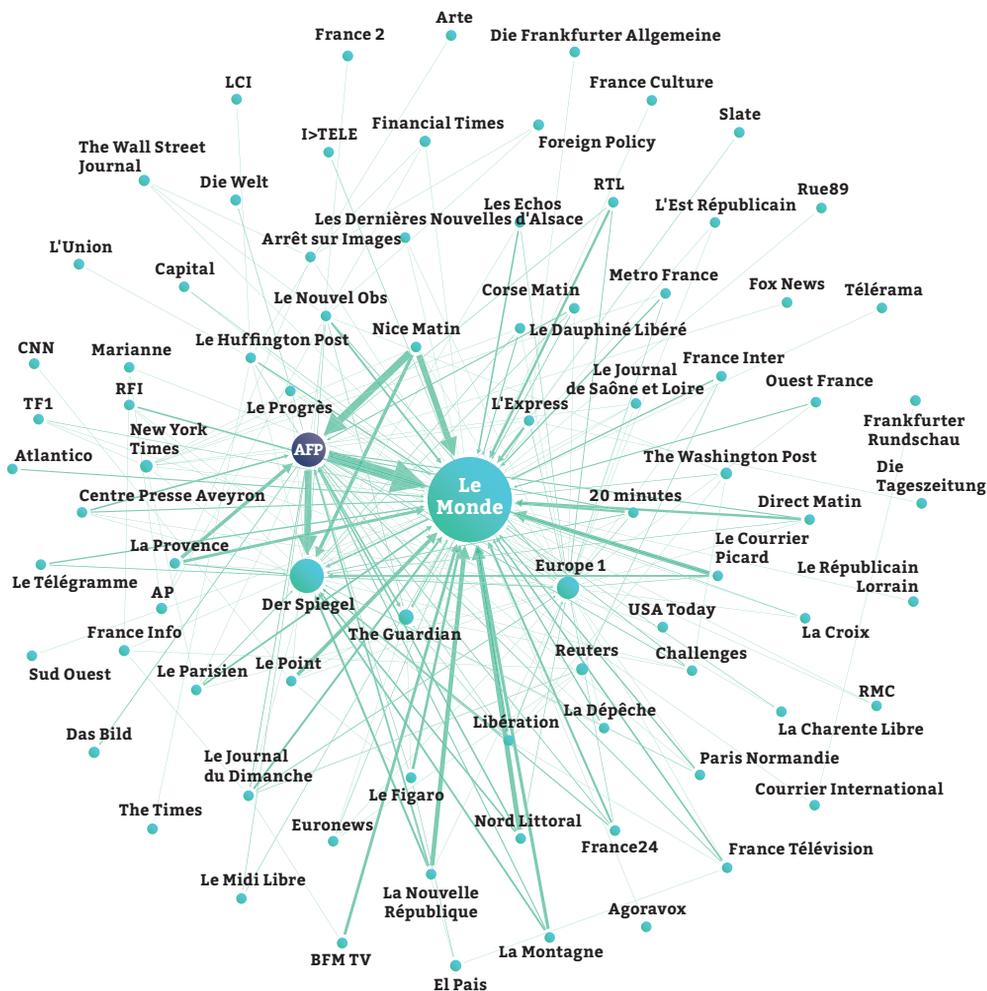


Pierre Elkabbach sur les écoutes, il les qualifie de « choquantes », propos qui seront ensuite repris par de nombreux autres médias<sup>1</sup>.

### *Une illustration du découpage en événements*

Cet exemple des écoutes de la NSA illustre bien la façon dont notre algorithme crée des événements et y classe les différents articles de manière cohérente. En effet, il identifie non pas un, mais trois événements liés aux écoutes le 21 octobre 2013. Après la révélation sur l'enregistrement des données, le deuxième événement – distinct du premier car il traite en fait d'un sujet différent – regroupe les articles couvrant la convocation par Laurent Fabius, alors ministre des Affaires Étrangères, de l'ambassadeur américain à Paris. Le troisième, enfin, traite de la convocation de l'ambassadeur américain par Mariano Rajoy, en Espagne.

<sup>1</sup> Les références à *Der Spiegel* et au *Guardian* s'expliquent par le fait que, comme nous l'avons noté plus haut, ces deux médias avaient précédemment évoqué une partie des documents de la NSA dévoilés par Snowden.



→ Figure 16

## LE MONDE, SNOWDEN ET LES ÉCOUTES DE LA NSA

La manière dont sont déterminés les seuils adéquats de proximité linguistique est donc cruciale. Si les seuils que nous avons fixés étaient tels que tous les articles publiés le 21 octobre et incluant les mots « NSA » ou « écoutes » étaient tous classés dans un même événement, alors beaucoup d'information serait perdue. La révélation des écoutes de la NSA est sans aucun doute l'histoire la plus importante, mais les convocations des ambassadeurs sont en eux-mêmes des événements, dont *Le Monde* n'a cette fois pas eu l'exclusivité.

## **Mediapart et l'enquête judiciaire sur le micro-parti de Marine Le Pen**

Quand on évoque un scoop en 2013 mis à jour par *Mediapart*, on pense automatiquement à la désormais célèbre « affaire Cahuzac ». Ce n'est pas l'exemple que nous avons choisi de détailler ici, car l'affaire a été révélée par *Mediapart* dès le 4 décembre 2012, dans un article de Fabrice Arfi intitulé « Le compte suisse du ministre du budget » qui malheureusement n'est pas inclus dans notre base de données (qui ne débute qu'au 1<sup>er</sup> janvier 2013). Cette révélation de *Mediapart* est toutefois elle-aussi un cas d'école, non seulement du fait de la qualité de l'enquête menée par le *pure player*, mais également des difficultés auxquelles celui-ci a dû faire face, immédiatement critiqué pour ne pas fournir suffisamment de preuves puis devant faire face quelques mois plus tard au prétendu « blanchiment » du ministre par la banque UBS<sup>1</sup>.

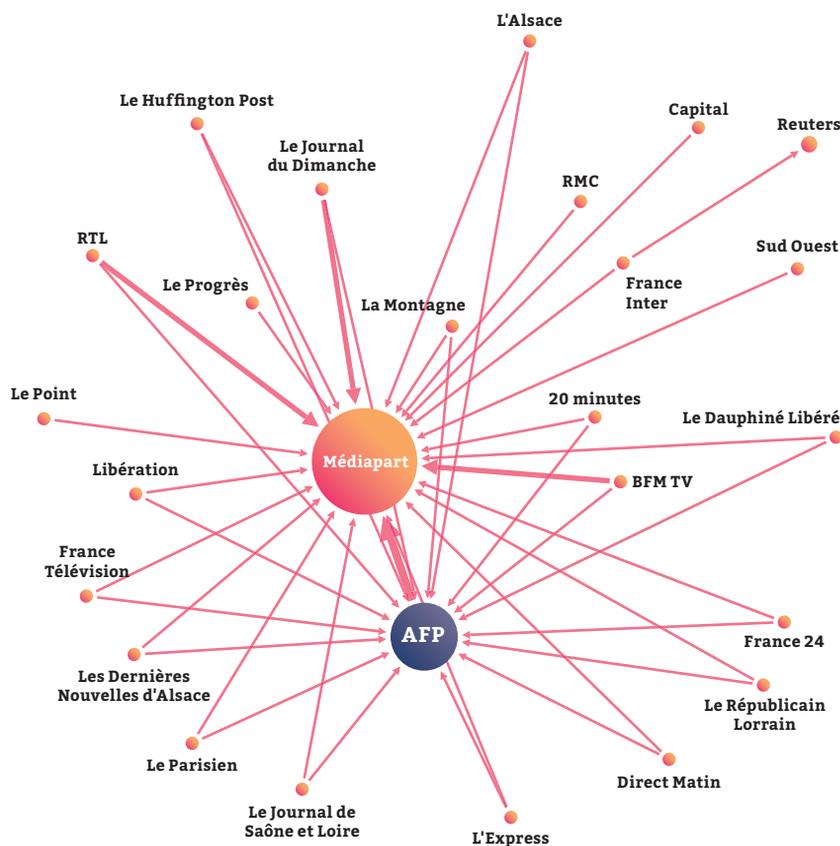
L'événement que nous avons choisi de détailler ici est d'une certaine façon moins « spectaculaire » mais plus représentatif de la nature des événements couverts chaque jour par les médias. Le 20 novembre 2013, *Mediapart* publie à 22h14 sur son site internet un article de 3 977 caractères selon lequel, « d'après nos informations », « la justice enquête sur le micro-parti de Marine Le Pen », Jeanne. L'enquête elle-même a en fait été ouverte dès le printemps 2013 par le parquet de Paris et confiée à la brigade financière, mais il faut attendre les informations de *Mediapart* pour que le grand public en soit informé.

De manière révélatrice, cette information de *Mediapart* n'est reprise par les autres médias que près de 13 heures plus tard, le lendemain, suite à la publication à 11h27 d'une dépêche AFP sur le sujet (« Une enquête judiciaire en cours sur le micro-parti de Marine Le Pen »). Il faut attendre la dépêche AFP pour que l'information de *Mediapart* soit publiée – cette fois rapidement – par les autres médias, avec un premier article mis en ligne à 12h02 sur le site de *L'Express*, suivi d'un autre par *20 Minutes* à 12h57, etc.

Au final, ce seront 29 articles sur cette enquête qui seront publiés le 21 novembre (le lendemain de la publication par *Mediapart*), dont trois dépêches de l'AFP. Mais la plupart de ces articles,

<sup>1</sup> Voir par exemple l'article du *Monde* résumant les principales étapes de l'affaire Cahuzac : <[www.lemonde.fr/politique/article/2012/12/10/si-vous-n-avez-rien-suivi-de-l-affaire-cahuzac\\_1804062\\_823448.html](http://www.lemonde.fr/politique/article/2012/12/10/si-vous-n-avez-rien-suivi-de-l-affaire-cahuzac_1804062_823448.html)> [consulté le 8 février 2016].

s'ils font référence à *Mediapart* (la **FIGURE 17** illustre la structure des citations), notent également que l'information de *Mediapart* a été confirmée par une source. Ainsi de *L'Express* : « Le parquet de Paris a ouvert en avril une enquête préliminaire, toujours en cours, sur le micro-parti de la présidente du FN, Marine Le Pen, a-t-on appris jeudi de source judiciaire, confirmant une information de *Mediapart* », ou de *20 Minutes*. Une telle confirmation n'avait pas semblé nécessaire dans le cadre des révélations du *Monde* sur les écoutes de la NSA.



→ Figure 17

## MEDIAPART ET L'ENQUÊTE JUDICIAIRE SUR LE MICRO-PARTI DE MARINE LE PEN

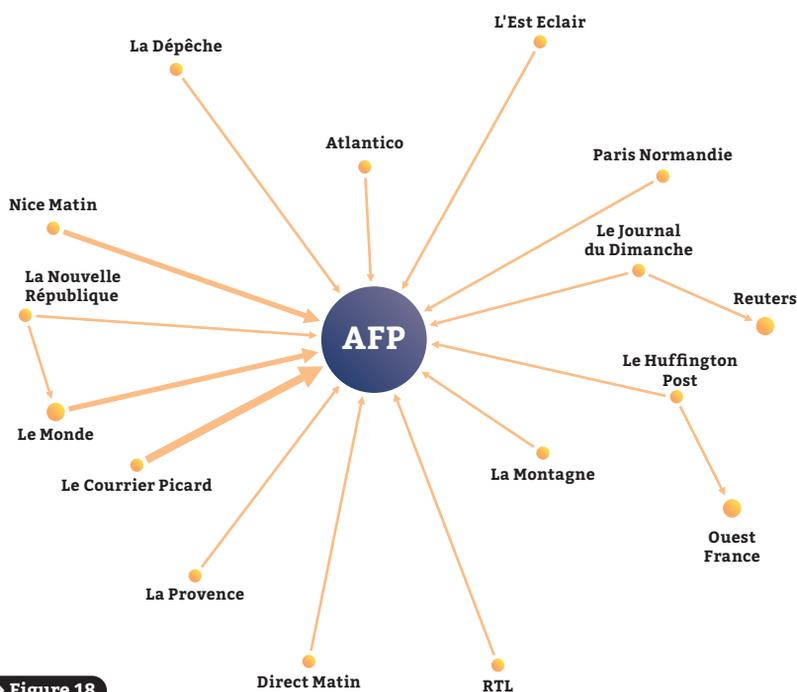
Ces reprises sont pour la plupart indiquée sous la forme « avec AFP », ce que l'on retrouve dans le taux de copie très élevé de ces articles par rapport à la dépêche initiale de l'agence (88% pour l'article de *L'Express*, 89% pour celui de *20 Minutes*). De manière intéressante, l'AFP, si elle reprend l'information de *Mediapart* (auquel elle fait référence), écrit une dépêche dont le verbatim diffère : cette dépêche est entièrement originale. Alors que 99% du texte de la dépêche seront repris par au moins un autre média, ce n'est le cas que de 11% de l'article de *Mediapart*, le premier pourtant à avoir révélé cette enquête. *Mediapart*, rappelons-le, est en 2013 l'un des seuls médias français à se trouver derrière un mur payant.

### **L'AFP annonce la candidature de François Fillon à la primaire**

Prenons un dernier exemple, cette fois révélé par l'AFP. C'est un événement « moyen » de notre corpus, représentatif de tous ces événements qui commencent par une dépêche AFP et donnent lieu à quelques reprises, voire à de légers développements.

Le 8 février 2013, l'AFP publie une dépêche selon laquelle François Fillon s'apprête à organiser un « grand meeting » le 26 février à la Mutualité de Paris, meeting au cours duquel « il devrait "annoncer sa candidature" à la primaire de son camp ».

Cet événement donne lieu à la publication de 35 articles dans la journée (dont quatre dépêches AFP), le premier média à mettre en ligne l'information étant *La Dépêche du Midi*, dans un article plus long que la dépêche (de l'AFP) mais reprenant l'essentiel de son contenu, et publié seulement quelques minutes plus tard. Après avoir noté que l'information a été confirmée par « plusieurs de ses proches » (la dépêche AFP précisait « selon plusieurs sources UMP »), *La Dépêche du Midi* cite l'agence comme la source de l'information à la fin de son article : « ... a-t-on dit à l'AFP. » L'article du *Parisien* publié quelques secondes plus tard – et qui reprend de même à son compte l'essentiel de la dépêche AFP – ne cite lui pas l'agence, pas plus d'ailleurs que *Le Journal du Dimanche* et le site internet d'Europe 1 dans les minutes qui suivent. Aucun de ces articles ne brille pourtant par son originalité : 13% pour le *JDD* et seulement 9% pour Europe 1.



→ Figure 18

### FRANÇOIS FILLON CANDIDAT À LA PRIMAIRE POUR 2017 : UNE INFORMATION DE L'AFP

Dans un second article, plus long (1980 caractères contre 587), publié quarante minutes plus tard sur son site, le *JDD* fait cependant clairement référence à l'AFP comme la source de l'information (« ... a révélé l'AFP vendredi »). Au total, ce sont 19 citations que l'AFP recevra dans le cadre de cet événement le 8 février, comme illustré sur la **FIGURE 18**. 98% du contenu de la première dépêche de l'Agence seront repris par au moins un autre média.



## CHAPITRE 3

### LES DÉTERMINANTS ÉCONOMIQUES DE LA PRODUCTION D'INFORMATION

Les différentes mesures de la production d'information sur Internet permettent de déterminer, pour chaque média, la quantité totale d'information produite ainsi que la quantité d'information originale, sa réactivité moyenne et sa tendance à être, ou non, le premier sur l'information.

Mais quels sont les facteurs permettant d'expliquer que certains médias produisent plus d'information que d'autres ?

#### Le coût de l'information : les journalistes

Une première réponse, qui peut paraître évidente mais qui est pourtant loin de faire l'unanimité, est la suivante : si certains médias produisent plus d'information que d'autres, et en particulier plus d'information *originale*, c'est parce qu'ils ont plus de journalistes.

Le nombre de journalistes a souvent été utilisé dans la littérature en économie et en sciences politiques comme une mesure indirecte de la qualité de l'information produite<sup>1</sup>. Cette approche par l'« input » n'est bien sûr pas entièrement satisfaisante, et de nombreuses critiques peuvent lui être adressées. En particulier, il est difficile de par-

---

1 Voir par exemple Hamilton, J. (2004). *All the News That's Fit to Shell : How the Market Transforms Information Into News*, Princeton University Press ; Berry, S. T. and Waldfogel, J. (2010). « Product Quality and Market Size », *Journal of Industrial Economics*, 58(1):1-31 ; Gentzkow, M. et Shapiro, J. M. (2010). « What Drives Media Slant? Evidence from US Daily Newspapers », *Econometrica*, 78(1) ; Fan, Y. (2013). « Ownership Consolidation and Product Characteristics : A Study of the US Daily Newspaper Market », *American Economic Review*, 103(5): 1598-1628 ; Cagé, J. (2014). « Media Competition, Information Provision and Political Participation », *op. cit.*

ler « des journalistes » en général et leur « productivité » est fonction de multiples facteurs, à commencer par leur expérience<sup>1</sup>. Elle est également fonction de leur environnement et dépend des moyens qui sont mis à leur disposition pour travailler.

De plus, il est difficile de parler de la « productivité » d'un journaliste. Mettre sur le même plan nombre de journalistes et qualité, n'est-ce pas confondre qualité et quantité? Certains articles peuvent parfois prendre des mois d'enquête et d'investigation de la part d'un groupe de journalistes qui ne produiront alors rien d'autre dans l'intervalle. Mais le ou les articles qui seront le fruit de ce long travail d'investigation (il suffit de penser aux Panama Papers dans l'actualité récente) doivent être considérés comme du travail journalistique de grande qualité. Les rédacteurs en chef ou les directeurs de rédaction écrivent parfois eux-mêmes relativement peu d'articles, mais leur travail d'orchestration est ce qui fait d'un média ce qu'il est.

Pourquoi alors faire le choix de cette approche par l'input pour mesurer la qualité de l'information? La réponse paraîtra sans doute à certains un peu décevante, mais si les chercheurs sont si nombreux à utiliser cette approche, c'est essentiellement parce que l'on ne sait pas faire mieux<sup>2</sup>.

De plus, nombre de journalistes et production d'information sont *de facto* fortement corrélés. Si l'approche par l'input est loin d'être parfaite, elle est donc néanmoins informative.

### *La taille des rédactions*

La taille de la rédaction varie énormément d'un média à l'autre<sup>3</sup>. En moyenne, la rédaction d'un média d'information comprenait 122

1 Patrick Champagne souligne à raison dans son dernier livre que « "le journaliste" n'existe pas. Ce qui existe, ce sont des journalistes qui occupent des positions déterminées [...] dans des journaux [...] qui constituent ce qu'il faut bien appeler un champ journalistique ». Champagne, P. (2016). *La Double Dépendance. Sur le journalisme*, Paris, Raisons d'Agir.

2 D'autres mesures sont parfois utilisées, comme les prix reçus par les journalistes (par exemple le Prix Pulitzer aux États-Unis), mais elles souffrent de leurs propres limites et permettent difficilement une approche globale. Les chercheurs pourraient souhaiter mesurer le degré d'information acquis par les citoyens à la lecture de tel ou tel article. Mais l'on voit que même ici les difficultés seraient nombreuses : ce qui est « informatif » du point de vue d'un lecteur peut facilement sembler « obscur » aux yeux d'un autre.

3 Les données que nous utilisons ici sur la taille des rédactions ont été collectées dans le cadre d'un travail de recherche sur l'évolution historique du nombre de journalistes en France et de leurs rémunérations : Cagé, J. (2016). « Payroll and Inequality within the Newsroom : Evidence from France, 1936-2016 », *op. cit.* Une de limites de ces données est que nous ne prenons en compte que les journalistes ayant leur carte de presse (y compris les pigistes). Or, notamment

journalistes en 2013, mais ces chiffres varient de 1 à près de 800. Ces variations ne s'expliquent pas par le support « hors ligne » des médias : la rédaction d'un journal papier compte en moyenne 123 journalistes, 161 pour les chaînes de télévision et 116 pour les radios.

Deux exceptions sont à noter : l'AFP, qui possède une rédaction de 780 journalistes titulaires de la carte de presse, et les *pure players*, dont les rédactions sont bien plus petites. Le cas des *pure players* suppose néanmoins d'insister sur deux précisions de taille.

Il est tout d'abord important de noter que les données sur la taille des rédactions sont des données sur les rédactions globales des médias. Si, lorsque les premiers sites internet, notamment des journaux, sont apparus, certains médias ont mis un point d'honneur à séparer leur rédaction web de leur rédaction papier<sup>1</sup>, cela n'est plus le cas aujourd'hui, bien au contraire. La très grande majorité des médias ont entièrement intégré leurs rédactions, ce qui était du point de vue des médias eux-mêmes sans doute la meilleure des décisions à prendre, mais vient néanmoins un peu compliquer le travail du chercheur... Car en intégrant leurs rédactions, les médias ont arrêté de reporter séparément le nombre de leurs journalistes travaillant sur le support digital. Il est loin d'être sûr que si ces données séparées avaient existé, ce sont celles qu'il aurait fallu utiliser, notamment parce que du fait de l'intégration des rédactions, ce qui est produit « hors ligne » est souvent publié « en ligne » et vice versa. Mais il est néanmoins important de souligner que dans le cas spécifique des *pure players*, il n'existe pas de monde « hors ligne », puisque la spécificité même des *pure players* est d'être un support qui n'existe qu'en ligne (que sur Internet). Cela peut en partie expliquer la faiblesse relative de la taille de leur rédaction.

Il est vrai que les données que nous présentons ici remontent à 2013. Les *pure players* étaient alors en France un phénomène relativement récent (*Rue89* a été lancé en 2007, *Mediapart* en 2008). Or il est difficile pour un nouveau média de se lancer avec une rédaction de taille importante ; cela suppose le plus souvent d'avoir déjà atteint un certain équilibre financier. La rédaction de *Mediapart* par

---

en presse quotidienne départementale et régionale, les journaux ont parfois recours à des correspondants, informateurs attirés – mais qui continuent à exercer leur métier propre – disséminés sur l'ensemble du territoire et qui alimentent en nouvelles les bureaux tenus par des journalistes professionnels. C'est le cas par exemple pour *Ouest France*, connu pour son réseau de correspondants et dont nous sous-estimons malheureusement ainsi l'investissement rédactionnel.

<sup>1</sup> On se souvient du temps où la rédaction du monde.fr et la rédaction du journal *Le Monde* ne partageaient même pas les mêmes locaux à Paris !

exemple a, depuis 2013, presque doublé de taille, au moment même où les médias « traditionnels » ne cessent de couper dans leurs effectifs. Le monde de 2017 diffère donc de ce point de vue de celui d'il y a quatre ans, et si nous devions mener à nouveau cette étude aujourd'hui, les *pure players* y joueraient sans aucun doute un rôle beaucoup plus important.

## Un détour par les robots

Il est difficile de réfléchir aujourd'hui au lien entre nombre de journalistes et production d'information sans évoquer une nouvelle forme de journalisme : le « *robot journalism* ». Les journalistes « robots » utilisent des algorithmes de traitement automatique du langage naturel et de curation de contenus<sup>1</sup>.

### *Le traitement automatique du langage naturel*

La rédaction automatique de texte est en effet un domaine dont l'étendue des possibles commence à peine à s'ouvrir à nous. En France, une entreprise comme Syllabs, qui travaille avec de nombreux médias « de référence », est en pleine expansion et propose de déléguer la rédaction de certains articles à des robots. Ainsi le site internet du *Monde* fait régulièrement appel à sa technologie pour couvrir les résultats électoraux.

Mais cela est loin d'être nouveau. Aux États-Unis, Associated Press, pionnière dans ce domaine, confie depuis 2012 la rédaction de comptes rendus sportifs à des algorithmes. Cette méthode est aujourd'hui de plus en plus utilisée pour le traitement de l'information économique, et en particulier les bilans d'entreprise.

Autre innovation, l'application Wibbitz – un programme d'intelligence artificielle – ne propose rien de moins que de... transformer tout article écrit en un résumé vidéo en moins d'une minute à l'aide de la technologie Text-to-Video<sup>2</sup>. Wibbitz a d'ailleurs annoncé en juillet 2016 un partenariat en France avec le groupe TF1, après

1 La curation de contenus est la pratique qui consiste à sélectionner, éditer et partager les contenus les plus pertinents publiés sur Internet sur un sujet donné.

2 La technologie Text-to-Video consiste à analyser le texte d'un article puis à trouver automatiquement des photos, des vidéos et des graphiques qui sont utilisés pour créer une vidéo sur le thème de l'article.

avoir établi un partenariat en juin avec Prisma Media (le plus grand éditeur de magazines en France). Pour TF1, il s'agit d'enrichir son offre vidéo... à faible coût : une rédaction « humaine » peut produire quelques dizaines, quelques centaines de vidéos au maximum par jour. Des logiciels Text-to-Video tels que celui de Wibbitz permettent d'en produire des milliers<sup>1</sup> !

On pourrait s'interroger sur l'utilité pour un éditeur de démultiplier à l'infini les vidéos disponibles en ligne. La réponse est simple : les tarifs des publicités pré-vidéos sont aujourd'hui beaucoup plus élevés que ceux des autres publicités en ligne. Ce à quoi il faut rajouter une demande croissante des réseaux sociaux, en particulier de Facebook, qui ont fait le pari d'une prédominance de la vidéo sur le texte dans le futur<sup>2</sup>.

#### *Avantages et limites des robots*

Que penser de cette nouvelle manière de « faire du journalisme » ? Doit-on s'en réjouir ou la craindre ? Répondre à ces questions dépasse de beaucoup l'objectif de ce livre, mais il est néanmoins utile, dans le contexte de l'étude de la production d'information sur Internet, d'apporter quelques pistes de réponse qui nous seront utiles pour la suite de l'analyse.

Si l'utilisation des robots se développe aussi vite, ce n'est pas (seulement) parce que les éditeurs voudraient réduire les coûts humains de la production d'information et remplacer à terme leurs journalistes par de l'intelligence artificielle. C'est aussi parce que l'humain a ses limites et que le recours aux robots permet de produire – rapidement – de l'information à une échelle qui n'était jusqu'à présent pas envisageable. Prenons un exemple révélateur, celui des élections. Pour des élections locales comme les élections municipales, une couverture exhaustive de l'actualité suppose de produire de l'information en France sur plus de 36 000 communes. Cela demande beaucoup de temps et plusieurs centaines de journalistes ne suffiraient pas à produire cette information en une seule nuit<sup>3</sup>. La délégation d'articles

1 On peut également mentionner Wochit, principal concurrent de Wibbitz, qui a signé en juin 2016 un partenariat avec Time Inc.

2 Ce qui suppose d'adapter le format des vidéos aux nouveaux modes de consommation. Franceinfo, la chaîne publique d'information en continu lancée en septembre 2016, a ainsi fait le choix de modules vidéo mêlant le texte et l'image et regardables sans son.

3 Le dispositif mis en place par Associated Press aux États-Unis pour l'élection présidentielle

à des robots le permet et c'est ainsi que *Le Monde* offre désormais à lire sur son site internet des résultats exhaustifs au cours de la soirée électorale.

La limite des algorithmes – qui est la raison pour laquelle ils ne pourront jamais prendre tout à fait la place des journalistes même s'il ne faut pas sous-estimer les progrès de l'intelligence artificielle –, c'est qu'ils reposent principalement sur la curation et le traitement de contenus déjà existants. Autrement dit, les robots peuvent produire de l'information, beaucoup d'information, mais relativement peu d'information véritablement *originale*<sup>1</sup>. Soulignons également que si la mise en forme automatique de tableaux de chiffres peut apporter une meilleure lisibilité au lecteur, elle est également un avantage non négligeable pour le média en vue de son indexation dans les moteurs de recherche. En effet ces derniers favorisent dans leur classement des sites web ceux qui offrent davantage de contenu original et correctement écrit. La capacité de production de ces robots vient donc interroger fortement ce qui serait une approche purement quantitative de l'information. Une quantité insoupçonnable d'« information » peut aujourd'hui être générée sans aucun effort, à partir d'un simple clic de souris<sup>2</sup>. D'où l'importance – c'est ce que nous faisons ici – de distinguer la masse d'information produite de l'information *originale*.

Outre ces robots, le recours au journalisme de données (*data journalism* ou *computational journalism*) permet également aux médias d'enquêter sur des sujets dont la complexité requiert le recours à des algorithmes et à l'exploitation de données massives. James Hamilton l'illustre fort bien dans son dernier livre au titre révélateur – les détectives de la démocratie – avec l'exemple du financement des

---

américaine (l'« Election Day ») est de ce point de vue extrêmement impressionnant avec plusieurs milliers de pigistes d'un jour qui appellent durant plusieurs heures chaque bureau électoral pour en faire remonter les résultats.

1 Soyons toutefois prudents ici puisque cela dépend de ce que l'on considère comme une information « originale ». D'une certaine manière, en traitant massivement des données et en produisant ainsi des statistiques, c'est une information « originale » que mettent à jour les robots ; sans l'aide des algorithmes de fouille de données, les 91 000 documents du WikiLeaks n'auraient jamais pu être exploités comme ils l'ont été, sans parler des 11,5 millions de documents derrière le leak des Panama Papers. Mais même dans ce cas, l'algorithme n'a de sens que par la médiation du journaliste. Ainsi, ce sont les journalistes des différents médias associés qui ont décidé ce qui devait être – au nom de l'information – ou ne pas être – au nom du respect de la vie privée – diffusé dans le cadre de cette investigation.

2 De nouveaux médias comme Brief.me ont au contraire fait le choix d'une curation et d'un traitement de l'information entièrement manuel, autrement dit de prendre le contre-pied des algorithmes.

campagnes électorales<sup>1</sup>. Les informations sur les dons reçus par les candidats et leurs dépenses sont depuis longtemps publiées sur le site de la *Federal Electoral Commission* (FEC) aux États-Unis ; mais le site de la FEC n'est pas très facile d'utilisation (et c'est une litote). Au contraire, l'interface de programmation (API) mise en place par ProPublica<sup>2</sup> permet d'étudier facilement et en temps réel les dons et les dépenses disponibles sur le site de la FEC, et le programme Bedfellows créé par Derek Willis sur le blog Upshot du *New York Times* d'explorer dans le détail les relations de long terme entre donateurs et bénéficiaires<sup>3</sup>. Pourquoi se priver de ces technologies ?

L'automatisation est un formidable progrès pour le journalisme si elle permet aux journalistes de se concentrer sur leur travail de journaliste, et d'éviter de dépenser du temps et de l'énergie dans des tâches répétitives pour lesquelles leur valeur ajoutée est faible. Il faut toutefois être prudent et éviter qu'elle ne prenne trop de place dans le futur. Générer automatiquement des articles analysant des bilans d'entreprise, cela a le mérite d'économiser du temps. Mais aucune intelligence artificielle n'aurait mis à jour à travers l'analyse automatique de ses bilans le scandale d'Enron.

L'exemple de la génération automatique de vidéos doit également nous rappeler que, pour que le citoyen soit bien informé, il doit aussi être « protégé » d'un déferlement infini d'informations. Car l'un des enjeux dans le monde d'aujourd'hui caractérisé par l'« infobésité » ou surcharge informationnelle est celui de l'attention, comme l'a fort bien montré Yves Citton<sup>4</sup>.

## **Production d'information et taille des rédactions : éléments chiffrés**

Le nombre de journalistes des médias d'information en France est-il corrélé à leur production d'information, et en particulier d'information originale ?

1 Hamilton, J. T. (2016). *Democracy's Detectives : The Economics of Investigative Journalism*, Harvard University Press.

2 ProPublica's Campaign Finance API.

3 Voir D. Willis, « New Software Helps Connect the Money Dots in Politics », 19 décembre 2014, en ligne : <[www.nytimes.com/2014/12/20/upshot/new-software-helps-connect-the-money-dots-in-politics.html](http://www.nytimes.com/2014/12/20/upshot/new-software-helps-connect-the-money-dots-in-politics.html)>

4 Citton, Y. (2014). *Pour une écologie de l'attention*, Couleur des idées, Seuil.

## Quelques biais potentiels

Les documents sur lesquels sont basées nos analyses ont été collectés automatiquement. Outre quelques problèmes techniques (par exemple à l'été 2013), nous tenons à souligner deux types de biais susceptibles d'apparaître dans notre corpus.

Une partie des indicateurs que nous calculons est liée à l'heure de publication des articles. Cette donnée est fournie par le média lui-même. En dehors de l'AFP, qui de par son modèle nous livre directement les dépêches, nos robots collectent les articles sur les sites des différents médias. Dans ce cas, l'heure de publication est extraite, soit du flux RSS, soit directement dans le contenu de la page web de l'article. Or, il arrive que cette heure de publication ne soit pas correcte. Nous avons identifié trois cas expliquant potentiellement une erreur sur la date de publication d'un article. Le plus souvent, il s'agit d'un décalage de quelques minutes qui peut s'expliquer par le fait que les serveurs ne sont pas tous correctement paramétrés. Ceci est assez peu probable tant les mécanismes pour garantir qu'un serveur est à l'heure sont standards, inclus dans tous les systèmes d'exploitation et largement éprouvés. Il faut donc plutôt y voir une stratégie d'antidatage des publications qui s'explique par la volonté d'apparaître en premier dans les index de Google et d'accroître ainsi le trafic sur son site. Nous avons également pu constater des erreurs manifestes chez certains médias, comme un décalage d'une heure apparaissant le jour du changement d'heure été/hiver, ce décalage pouvant parfois n'être corrigé que quelques mois plus tard. Enfin, il arrive que l'information ne soit tout simplement pas disponible et qu'un site ne fasse apparaître que la date de publication sans préciser l'heure. Pour tous ces cas, nous avons corrigé dans la mesure du possible les données.

Le second biais potentiel vient de l'extraction de l'article depuis les pages web. Une page HTML est un format dédié à la mise en forme de son contenu par un navigateur web, il n'est pas initialement prévu pour permettre l'échange de

données structurées entre ordinateurs. Il peut ainsi arriver que les algorithmes chargés d'extraire le contenu et les métadonnées liées à un article commettent des erreurs. Typiquement, il peut arriver que l'on conserve du texte qui n'est pas directement lié à l'article tel que les titres des autres actualités du jour. Ce problème est maintenant résolu avec les évolutions du HTML et l'inclusion de métadonnées directement au sein des pages web, ce n'était pas encore complètement le cas en 2013 au moment où nous avons constitué notre corpus.

Ces deux types de biais peuvent légèrement influencer sur les résultats individuels d'un média et il est important de bien les avoir en tête avant de lire la suite de ce chapitre. En revanche, pris globalement, ils n'ont pas d'incidence significative sur les résultats agrégés que nous avons déjà présentés.

### *Une première approche*

La **FIGURE 19 (A à F)** présente les résultats que nous avons obtenus de manière visuelle. Sur l'axe des abscisses, nous avons reporté le nombre de journalistes des différentes rédactions ; sur l'axe des ordonnées, la production d'information *agrégée* sur l'année 2013 des différents médias<sup>1</sup>. Pour mesurer la quantité d'information produite, nous avons fait le choix d'utiliser le nombre total de caractères plutôt que le nombre total d'articles, la taille des articles pouvant varier très fortement.

Les différentes mesures de la production d'information que nous considérons ici sont les suivantes : la production d'information totale, classée ou non dans des événements ; le nombre d'événements couverts ; la production d'information *originale* (à l'intérieur des événements) ; et le nombre d'événements pour lequel le média a été *news breaker*.

<sup>1</sup> La production agrégée permet une meilleure visualisation de ces corrélations. Les résultats que nous présentons par la suite utilisent des variations beaucoup plus fines, en étudiant simultanément l'audience et la production *quotidienne* des différents médias.



→ Figure 19 B

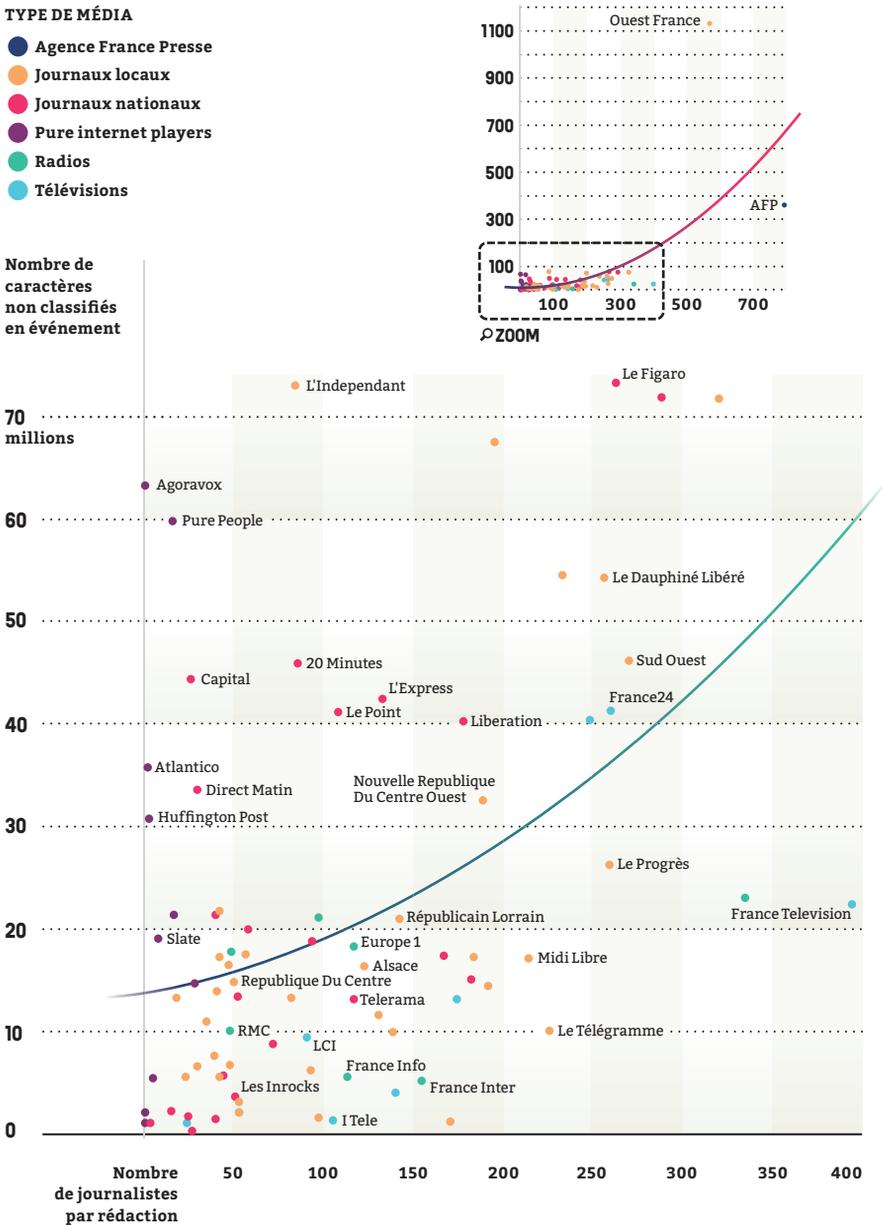
## PRODUCTION D'INFORMATION ET TAILLE DES RÉDACTIONS

Production non classifiée en événement et taille des rédactions

### TYPE DE MÉDIA

- Agence France Presse
- Journaux locaux
- Journaux nationaux
- Pure internet players
- Radios
- Télévisions

Nombre de caractères non classifiés en événement



→ Figure 19 C

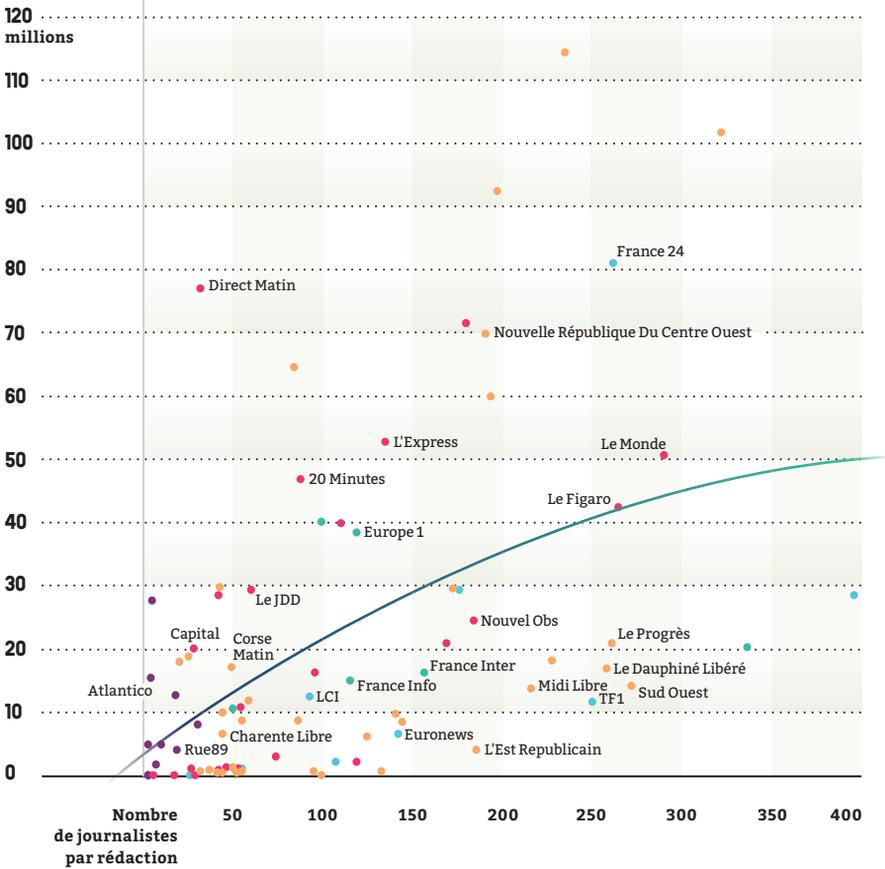
## PRODUCTION D'INFORMATION ET TAILLE DES RÉDACTIONS

Production classifiée en événement et taille des rédactions

### TYPE DE MÉDIA

- Agence France Presse
- Journaux locaux
- Journaux nationaux
- Pure internet players
- Radios
- Télévisions

Nombre de caractères classifiés en événement



→ Figure 19 D

## PRODUCTION D'INFORMATION ET TAILLE DES RÉDACTIONS

Événements couverts et taille des rédactions

### TYPE DE MÉDIA

- Agence France Presse
- Journaux locaux
- Journaux nationaux
- Pure internet players
- Radios
- Télévisions

Nombre d'événements couverts

20000

18000

16000

14000

12000

10000

8000

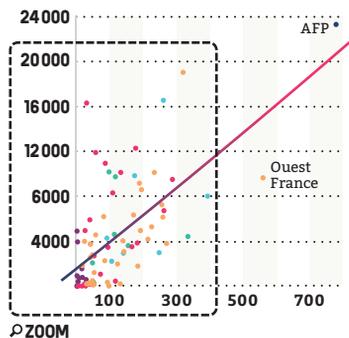
6000

4000

2000

0

Nombre de journalistes par rédaction



50 100 150 200 250 300 350 400

Direct Matin

France 24

Le Monde

Liberation

Le JDD

20 Minutes

L'Express

Le Point

Le Dauphiné Libéré

Le Figaro

Le Progrès

Capital

France Info

LCI

La Dépêche Du Midi

Le Télégramme

Midi Libre

Bien Public

Challenges

RMC

La Croix

Republicain Lorrain

Sud Ouest

Pure People

Alsace

L'Est Republicain

Courrier International

→ Figure 19 E

## PRODUCTION D'INFORMATION ET TAILLE DES RÉDACTIONS

Contenu original et taille des rédactions

### TYPE DE MÉDIA

- Agence France Presse
- Journaux locaux
- Journaux nationaux
- Pure internet players
- Radios
- Télévisions

Nombre de caractères originaux

35 millions

30

25

20

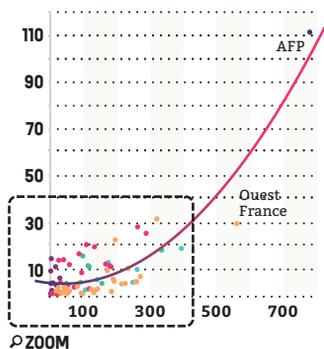
15

10

5

0

Nombre de journalistes par rédaction



35 millions

30

25

20

15

10

5

0

0 50 100 150 200 250 300 350 400

Le Figaro

Le Monde

Le Point

L'Express

20 Minutes

Metrofrance

Le JDD

Pure People

Atlantico

La Croix

Mediapart

Rue89

Challenges

LCI

Europe 1

Nouvel Obs

France Inter

Euronews

Nouvelle République Du Centre Ouest

France 24

France Télévision

Sud Ouest

Le Dauphiné Libéré

L'Est Republicain

Ouest France

AFP

→ Figure 19 F

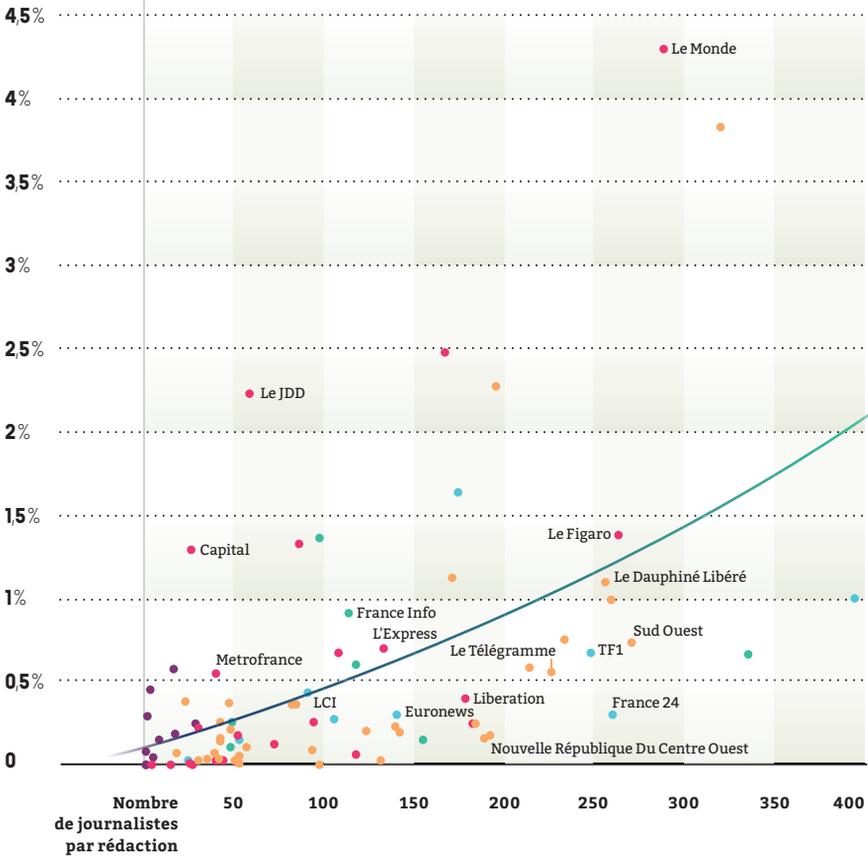
## PRODUCTION D'INFORMATION ET TAILLE DES RÉDACTIONS

Breaking news et taille des rédactions

### TYPE DE MÉDIA

- Agence France Presse
- Journaux locaux
- Journaux nationaux
- Pure internet players
- Radios
- Télévisions

Pourcentage de breaking news



Quelle que soit la mesure de production d'information que l'on considère, un acteur se distingue clairement des autres : l'AFP. Avec une rédaction de plus de 750 journalistes, l'AFP produit non seulement la plus grande quantité de contenu – original ou non – mais est également le premier *news breaker* en France. L'AFP est à l'origine de la moitié des événements et couvre 93% des événements. Bien sûr, cela est en partie lié au modèle économique spécifique de l'AFP, qui est un modèle BtoB.

Avec ou sans l'AFP, la production d'information – une fois encore, quelle que soit la mesure considérée – est très fortement corrélée à la taille des rédactions. La quantité d'information totale augmente avec le nombre de journalistes, tout comme le nombre d'événements couverts, la quantité d'information originale et la probabilité pour le média d'être *news breaker*. L'information sur Internet provient en grande partie des médias dont les rédactions sont les plus grandes.

De plus, à l'exception des *pure players* dont nous avons souligné la spécificité, les médias ne diffèrent pas « en ligne » en fonction de leurs supports « hors ligne ». Cette réalité, qu'il est important de rappeler, peut s'expliquer en partie par le fait que sur Internet, les médias ne diffèrent que relativement peu. À l'exception des renvois vers le live des télévisions/radios sur certains sites, rien ne différencie le site internet d'un journal, d'une télévision ou d'une radio.

Si différence il y a, elle est bien davantage entre médias nationaux et médias locaux. À taille de rédaction similaire, médias nationaux et locaux produisent une quantité d'information relativement semblable. Mais pour les médias locaux, la proportion d'articles à être classés dans des événements est plus faible (et le nombre d'événements traités l'est également). Cela est en partie dû à la manière dont nous avons défini les événements. Pour qu'il y ait événement, il faut qu'au moins deux médias le traitent, ce qui exclut de fait les événements locaux qui ne font pas l'objet d'un traitement en dehors des médias locaux.

### *Estimations économétriques*

Ce livre est en grande partie le fruit d'un travail de recherche combinant les outils des sciences informatiques et ceux de l'économie. En particulier, nous avons eu recours aux outils de l'économétrie pour estimer quantitativement les déterminants de la production d'information (**ENCADRÉ**). Nous ne pouvons entrer ici dans tous les

détails de l'évaluation économétrique : nous invitons donc tous ceux que ces détails intéressent à se reporter aux articles de recherche en anglais sur lequel ce livre se base et qui sont disponibles librement en ligne sur le site internet des auteurs<sup>1</sup>.

### **L'approche économétrique : quelques éléments de compréhension**

L'ambition de l'économétrie est de mettre en évidence empiriquement et de quantifier des relations causales entre phénomènes économiques. Il peut s'agir par exemple de quantifier l'impact causal d'un journaliste supplémentaire sur la production d'information originale par un média. La production d'information originale est la « variable expliquée » ; le nombre de journalistes la « variable explicative ». Imaginons que l'on sache très précisément – ce qui n'est pas le cas puisque tout l'exercice consiste justement à l'estimer – que cet impact soit égal à  $x$  : autrement dit, un journaliste supplémentaire augmente chaque jour en moyenne de  $x$  articles la production d'information originale.  $x$  est le paramètre que l'on cherche à estimer<sup>2</sup>.

Nous utilisons dans le cadre de cette recherche les méthodes d'estimation. L'estimation consiste à quantifier un paramètre théorique, tel qu'il est défini pour une population, à partir de l'observation d'une partie de la population, appelée échantillon.

Pour être valide, une procédure d'estimation doit être « sans biais » : les estimations obtenues à partir d'un ou plusieurs échantillons doivent permettre d'obtenir la « vraie » valeur moyenne (ici  $x$ ). Elle doit également être convergente : en augmentant la taille de l'échantillon – par exemple en ayant

.....  
<sup>2</sup> Les méthodes économétriques sont nombreuses et sont décrites de manière très accessible dans l'excellent petit livre de Luc Behaghel, *Lire l'économétrie* (Paris, La Découverte, 2006), sur lequel nous nous appuyons dans cet encadré.

<sup>1</sup> Voir <<https://sites.google.com/site/juliacagehomepage/research>>.

des informations sur l'ensemble des médias européens, de leur production d'information et de leurs journalistes – elle doit nous permettre de nous approcher avec un risque d'erreur de plus en plus petit de la vraie valeur ( $x$ ). Il est ainsi essentiel d'estimer la précision de chaque estimation. La précision de l'estimation nous permet de connaître avec quel risque d'erreur un paramètre peut être situé dans une fourchette donnée. Est qualifiée de « statistiquement significative » la valeur estimée d'un paramètre dont le risque d'erreur est suffisamment faible.

Il est cependant important de noter qu'un paramètre estimé peut être tout à la fois statistiquement significatif et quantitativement faible. C'est ce que nous verrons lorsque nous estimerons l'effet de la production d'information originale sur l'audience des médias. Significativité statistique et significativité quantitative sont deux dimensions distinctes.

Le nombre de journalistes travaillant pour un média peut-il à lui seul expliquer le nombre d'articles publiés chaque jour par ce média ? Il est évident que non et c'est pourquoi l'économètre a recours à la régression multiple (et à la technique associée des moindres carrés ordinaires). Autrement dit, il cherche à expliquer une variable par plusieurs variables explicatives. C'est pourquoi tout au long de ce livre nous utiliserons systématiquement un grand nombre de variables explicatives, parfois également appelées variables de contrôle.

Notons pour finir que ces estimations, afin de déterminer des relations causales, se font « toutes choses égales par ailleurs ». La régression multiple permet autant que possible de prendre en compte l'impact simultané de plusieurs variables explicatives sur une variable expliquée, afin de limiter le biais que provoquerait leur omission et de faire ainsi un pas important en direction de la causalité.

Nous avons fait le choix de ne reporter que le résultat des estimations et non les estimations elles-mêmes. Avant de présenter le résultat de ces estimations, il est important de souligner une chose.

Quand, dans le chapitre 5, nous présenterons le lien entre la production d'information par les médias et la variation de leurs parts d'audience au niveau quotidien, les résultats présentés seront des résultats causaux. L'estimation causale repose sur l'utilisation des variations quotidiennes. Une telle estimation n'est pas possible dans le cas des déterminants de la production d'information, les données sur la taille des rédactions des médias n'étant disponibles qu'au niveau annuel. Les résultats présentés ci-dessous ne sont donc que des corrélations.

En régressant les différentes mesures de production d'information sur le nombre de journalistes (ainsi qu'un certain nombre de contrôles<sup>1</sup>), nous obtenons une corrélation positive et statistiquement significative entre le nombre de journalistes et la production d'information, consistante avec l'évidence graphique présentée plus haut.

Cette corrélation n'est pas seulement statistiquement significative mais également économiquement significative, et elle est plus forte pour la production d'informations classées dans des événements que pour la production d'information contextuelle ou les articles d'actualité locale non classés dans des événements. En moyenne annuelle, une augmentation de 1% du nombre de journalistes travaillant dans une rédaction augmente le contenu non classé dans des événements de 0,78%, le contenu classé de 1,29% et le contenu original de 1,20%. De plus, une augmentation du nombre de journalistes de 1% augmente le nombre de *breaking news* d'un peu plus d'1%.

Si l'on réfléchit en termes de nombre de journalistes, nos résultats montrent qu'un journaliste supplémentaire augmente en moyenne de 28 articles *entièrement originaux* par an la production d'information classée dans des événements d'un média<sup>2</sup>. La différence entre le nombre de journalistes du *Monde* et le nombre de journalistes de *Libération* possédant la carte de presse en 2013 – soit 110 cartes de presse – permet à elle-seule d'expliquer près de la moitié de la différence quantitative de production d'information originale entre ces deux médias (environ 12 millions de caractères originaux pour *Libération* en 2013 contre près de 26 millions pour *Le Monde*).

Cette augmentation de la production provient-elle du fait que les médias aux rédactions les plus grandes couvrent plus d'événements, et/ou du fait qu'ils publient plus d'articles (ou des articles plus longs) à l'intérieur des événements qu'ils couvrent? Pour répondre à cette

1 Modèle de régression linéaire utilisant la méthode des moindres carrés.

2 Sur la base d'une longueur moyenne de 1 865 caractères par article.

question, nous avons étudié la production d'information par les médias à l'intérieur de chaque événement<sup>1</sup>. Si la production d'information à l'intérieur des événements augmente avec le nombre de journalistes, la plus grande partie de l'effet provient du plus grand nombre d'événements couverts par les médias ayant plus de journalistes. Autrement dit, les plus grandes rédactions couvrent plus d'événements ; à l'intérieur de chaque événement, elles produisent également plus d'information, et en particulier plus d'information originale, mais cet effet est quantitativement moins important.

Bien sûr, il faut être prudent sur l'interprétation de ces chiffres et une approche purement quantitative – et productiviste – ne saurait être entièrement satisfaisante. Mais ces résultats n'en sont pas moins inquiétants dans ce qu'ils nous disent des conséquences à venir de la baisse actuelle des tailles de rédaction. Moins de journalistes demain, cela voudra dire moins d'information originale produite, et moins de *breaking news*. Quantitativement, la masse totale d'information donnée à lire aux citoyens n'est pas près de diminuer, avec la multiplication des blogs, des réseaux sociaux et autres agrégateurs, mais qualitativement, cette information risque de s'homogénéiser et surtout, des « événements » (par exemple des cas de corruption) risquent d'échapper à l'attention des journalistes et de ne jamais devenir des « événements médiatiques » donnés à lire aux citoyens pour éclairer leurs votes<sup>2</sup>. Or il faut préserver à tout prix la production d'une information de qualité.

### *Prochaines étapes : investigation et réseaux sociaux*

Les travaux de recherche sur lesquels se base ce livre souffrent de plusieurs limites, que nous avons soulignées. À l'exception du travail que nous avons mené sur leur thématique, nous avons par exemple eu tendance à traiter l'ensemble des événements de manière indiffé-

1 L'estimation au niveau des événements nous permet également de contrôler pour un plus grand nombre de variables explicatives, comme le sujet de l'événement ou son importance.

2 Dans son dernier livre, James Hamilton documente avec précision l'effondrement de l'investigation journalistique au niveau local aux États-Unis. Or, cet effondrement est en grande partie dû à la fermeture de nombreux journaux quotidiens locaux, ainsi qu'aux moindres ressources dont disposent ceux qui ont survécu. Hamilton, J. T. (2016). *Democracy's Detectives, op. cit.* Cet effondrement n'est de plus sans doute pas étranger au fait que les médias américains n'aient pas vu venir l'élection de Donald Trump en novembre 2016. Au-delà de l'entre soi d'une certaine élite journalistique, cette « myopie » s'explique également par la disparition d'un journalisme de proximité que l'on ne peut que déplorer. Espérons que nous saurons en tirer les leçons, des deux côtés de l'Atlantique.

renciée. En particulier, les algorithmes que nous avons développés ne nous ont pas permis d'identifier, pour chaque événement, d'où le média *news breaker* tirait son information. Or, la quantité de travail nécessaire au journaliste n'est pas la même dans le cas où l'information provient d'une tierce partie, par exemple des déclarations d'un homme politique dans le cadre d'une conférence de presse, que lorsqu'il s'agit d'un scoop révélé après des mois d'investigation.

Une telle distinction ne peut aujourd'hui n'être que le fruit d'un travail de codage manuel de l'ensemble des événements, montagne quand l'on sait que nous en avons identifié près de 25 000 pour la seule année 2013. Nous travaillerons dans le futur à automatiser cette analyse. Certes, les résultats de la littérature existante laissent à penser que l'investigation représente tout au plus 1% des articles publiés<sup>1</sup>, mais le journalisme d'investigation est au cœur de l'information comme garante du bon fonctionnement de nos démocraties. Il serait ainsi intéressant d'étudier dans quelle mesure les différents phénomènes que nous avons analysés – vitesse de propagation de l'information, importance du copié-collé... – sont identiques lorsque l'on se concentre sur cette forme particulière de production médiatique. Car l'investigation, plus encore que toutes les autres formes d'information, se caractérise tout à la fois par un coût de production extrêmement élevé et des bénéfices qui le sont encore plus pour l'ensemble de la société. Mais par forcément pour le média qui en a porté le coût.

Une autre des limites de nos travaux est le fait que notre corpus de médias ne comprend ni les agrégateurs, ni les réseaux sociaux, qui prennent aujourd'hui une importance croissante dans la propagation de l'information sur Internet. Cette importance est telle que certains affirment même aujourd'hui que le rôle des médias ne serait plus d'annoncer les informations mais de les confirmer. Ainsi selon Nicolas Vanderbiest, chercheur qui a analysé le parcours de l'information lors de l'attentat de Nice en juillet 2016, ce ne sont pas les réseaux sociaux mais les médias traditionnels et en particulier les chaînes d'information en continu qui ont propagé des rumeurs<sup>2</sup>. La prochaine

<sup>1</sup> Dans le cadre de l'analyse des unes de trois journaux américains – le *Milwaukee Journal Sentinel*, le *New York Times* et le *Washington Post*, Katherine Fink et Michael Schudson ont ainsi montré que les articles d'investigation ne représentaient aujourd'hui que 1% des articles mis en une. Fink, K. et Schudson, M. (2014). « The Rise of Contextual Journalism, 1950s-2000s », *Journalism*, 15(1): 3-20.

<sup>2</sup> Voir par exemple son blog « Reputation Lab ».

étape de nos travaux consistera à introduire les réseaux sociaux dans notre analyse de la propagation de l'information en ligne.

Nous espérons également dans le futur avoir l'opportunité de travailler directement et plus étroitement avec les éditeurs eux-mêmes. L'obtention de données de première main est essentielle pour les chercheurs pour pouvoir mener à bien des analyses empiriques de qualité.

## CHAPITRE 4

### LES AGENCES DE PRESSE : DES ACTEURS À PART

Les agences de presse jouent un rôle à part dans le paysage médiatique du fait de la quantité d'information qu'elles produisent chaque jour. De plus, le modèle économique spécifique de la plupart des agences – un modèle de BtoB, c'est-à-dire business-to-business, l'AFP s'adressant aux autres médias et non aux particuliers – rend intéressant leur étude pour qui réfléchit à des modèles alternatifs pour les autres médias<sup>1</sup>.

D'après la définition donnée par l'UNESCO en 1953 (*Les agences télégraphiques d'information*), une agence d'information a pour mission « de rechercher des nouvelles et d'une façon générale des documents d'actualité ayant exclusivement pour objet l'expression ou la représentation des faits et de les distribuer à un ensemble d'entreprises d'information et exceptionnellement à des particuliers en vue de leur assurer, contre paiement d'une redevance et dans des conditions conformes aux lois et usages du commerce, un service d'information aussi complet et impartial que possible ».

La Fédération Française des Agences de Presse (FFAP) compte aujourd'hui près de 100 agences de presse, représentant plus de 560 millions d'euros de chiffre d'affaires, soit près de 80% du chiffre d'affaires de la branche<sup>2</sup>. Notre corpus comprend l'ensemble des dépêches en français produites en 2013 par l'AFP<sup>3</sup>.

Pourquoi l'AFP? Premièrement, l'AFP était un partenaire du projet OTMedia et, à ce titre, a apporté ses ressources au projet,

---

1 Nous discuterons en particulier dans le chapitre consacré aux solutions de l'importance croissante de la syndication de contenus.

2 D'après les chiffres communiqués sur le site internet de la FFAP. La FFAP existe depuis 1945.

3 Soit un peu moins du quart de la production totale de l'AFP.

à savoir l'intégralité de ses dépêches<sup>1</sup>. Deuxièmement, l'AFP est à la fois la principale agence de presse française, et la troisième plus grosse agence de presse au monde. L'AFP n'est pas en effet – comme DPA en Allemagne, EFE en Espagne ou l'ANSA en Italie, pour n'en citer que trois – une agence de presse d'envergure nationale, mais bien – aux côtés d'Associated Press et de Reuters – une agence d'information généraliste à rayonnement mondial. Elle est présente de manière permanente dans près de 190 pays ou territoires à travers 260 points de présence continue. S'il avait fallu n'en choisir qu'une, en dehors de toute contrainte de données, nous aurions de toute évidence choisi d'utiliser les données de l'AFP qui représente à elle-seule plus de la moitié du chiffre d'affaires de la FFAP. C'est enfin une agence de presse généraliste, ce qui nous intéresse particulièrement ici dans le cadre de l'étude des médias d'information politique et générale.

### Une production d'information « à part »

Sur l'année 2013, notre base de données référence plus de deux millions et demi d'articles qui ont été produits en ligne par les médias d'information. L'AFP à elle-seule représente 12% de ces documents<sup>2</sup>. L'AFP est ainsi quantitativement l'un des principaux producteurs d'information en France, seulement devancée par *Ouest France* et ses très nombreuses éditions locales. Mais ses dépêches ne sont pas mises en ligne et donc non accessibles directement au grand public.

De plus, l'AFP couvre la quasi-totalité des événements que nous avons mis à jour pour 2013, puisqu'elle apparaît dans 93% d'entre eux (alors qu'en moyenne les autres médias n'apparaissent que dans 16% des événements<sup>3</sup>). Ce à quoi il faut ajouter qu'elle est le prin-

1 Nous espérons que de telles collaborations pourront, à l'avenir, s'établir avec d'autres agences de presse ayant une production francophone.

2 Les ordres de grandeur sont similaires si l'on regarde le nombre de caractères plutôt que le nombre d'articles.

3 Précisons que cela n'est pas par construction puisque nous n'avons pas défini les événements à partir des dépêches AFP, mais à partir de l'algorithme de détection d'événements décrit dans le chapitre 1. Mais la couverture « complète » de l'actualité est partie prenante du statut de l'AFP, qui précise que l'Agence est tenue de rechercher en France et à l'étranger « les éléments d'une information complète et objective ». Autrement dit, les fils d'informations de l'AFP doivent couvrir l'ensemble de l'actualité dans toutes ses composantes.

cial *news breaker* puisqu'elle est la première à couvrir près de la moitié des événements.

Cette production d'information très forte quand on se focalise sur les événements se fait-elle au détriment de la production d'une information moins « chaude », avec moins d'articles non classés dans des événements ou « contextuels » ? On pourrait le penser mais cela n'est en fait pas le cas. En moyenne sur notre corpus, 40% de l'information produite est classée dans des événements ; c'est le cas de 50% des dépêches produites par l'AFP. En d'autres termes l'AFP, malgré sa focalisation très forte sur l'actualité, publie également un grand nombre d'articles qui ne sont pas classés dans des événements.

Comment l'interpréter ? D'une part, une partie de ces articles relève de la production contextuelle et il est donc normal qu'ils ne rentrent pas dans des événements. C'est par exemple le cas d'une dépêche de 1 404 caractères publiée le 22 juillet 2013 par l'AFP sur « l'ordre de succession au trône britannique », ou encore d'une dépêche du 24 avril 2013 annonçant la mort – et reprenant la vie – de George Jones, chanteur américain de country.

D'autre part, un certain nombre des dépêches de l'AFP, bien que couvrant effectivement des événements, ne sont jamais reprises<sup>1</sup>. C'est le cas par exemple d'une partie de l'actualité internationale qui ne fait pas l'objet d'une couverture par les médias français, comme la dénonciation d'une persécution politique par l'homme le plus riche d'Équateur – Alvaro Noboa – auprès de la Commission interaméricaine des droits de l'homme (CIDH) en novembre 2013 (dépêche AFP de 1 405 caractères datée du 26 novembre et jamais reprise). C'est le cas également parfois de l'actualité française, sans doute jugée « trop technique » ou « pas assez intéressante » pour donner lieu à des reprises par les sites internet des médias d'information généralistes qui composent notre corpus. On peut citer par exemple le rachat par les fonds d'investissement Apax Partners et Altamir des écoles privées du groupe d'enseignement supérieur IN-SEEC (dépêche AFP du 28 octobre 2013). C'est le cas enfin d'une partie de l'actualité sportive. Les médias d'information politique et générale (rappelons ici que des médias plus spécialisés comme *L'Équipe* n'entrent pas dans notre échantillon) tendent en effet à ne couvrir que les sports les plus populaires.

<sup>1</sup> Et donc, du fait de notre définition d'un événement – au moins deux médias et 10 articles – n'entrent pas dans des événements.

## La plus grosse rédaction de France

Comment expliquer la quantité très élevée d'information produite quotidiennement par l'AFP? Tout d'abord, par la taille de sa rédaction qui comptait 780 cartes de presse en 2013<sup>1</sup>, à quoi il faut rajouter des rendements d'échelle qui semblent plus fortement croissants pour l'AFP que pour les autres rédactions. Si l'on prend la relation moyenne entre nombre de journalistes d'une part et nombre d'articles publiés d'autre part, les journalistes de l'AFP semblent en effet relativement plus « productifs » que ceux des principaux autres médias<sup>2</sup>.

Gardons néanmoins à l'esprit que les 780 cartes de presse en 2013 couvrent uniquement les journalistes sous contrat français, soit environ la moitié des effectifs journalistiques de l'AFP. En comparaison internationale, la taille de l'AFP est relativement réduite, puisque AP comme Reuters comptent 2 500 journalistes – et que Chine Nouvelle en déclare 6 000!

### *Les autres facteurs de la réussite de l'AFP*

Au-delà de l'importance de ses effectifs, d'autres facteurs peuvent expliquer l'importance de la production de l'AFP. D'une part, notamment en ce qui concerne les *breaking news*, l'AFP semble plus réactive que les autres médias dans la reprise d'informations publiées par les médias étrangers et les autres agences. D'autre part, du fait de son modèle économique en BtoB, l'AFP se doit d'être

1 Pour un total de 1 360 équivalents temps plein (ETP) sous contrat français, soit un ratio journalistes sur total des employés d'environ 57%, un peu supérieur à la moyenne des autres médias. Le nombre d'ETP hors de France est légèrement supérieur au nombre d'ETP en France.

2 Mais cela est à pondérer en partie par le fort taux de copie « interne » à l'AFP. Le système des dépêches est en effet tel que l'AFP en publie plusieurs d'affilée sur un même sujet, de taille et avec des degrés de précision différents. De plus, la veille internationale historiquement réalisée par l'AFP grâce à son réseau de bureaux à l'étranger – et des contrats d'échange de contenus avec d'autres agences nationales – l'amène à être le *news breaker* français sur des nouvelles internationales dont elle n'est pas la source originale. Même si ces dépêches nécessitent traduction et réécriture – notons d'ailleurs qu'historiquement à l'origine de l'Agence Havas, futur AFP, se trouve le bureau de traduction de Charles-Louis Havas –, elles n'en sont pas pour autant des créations de contenu original. Toutefois, si cela revient à surestimer la production de contenu par l'AFP, nous la sous-estimons en revanche par la prise en compte uniquement des dépêches en français qui représentent aujourd'hui moins du quart des dépêches (en français mais également en anglais, en espagnol, en portugais, en allemand ainsi qu'en arabe) publiées chaque jour par l'AFP.

la première sur l'information. Cela apparaît clairement lorsque l'on s'intéresse aux citations reçues par les médias. Sur les plus de 460 000 citations que nous avons identifiées pour l'année 2013, près de 50% sont des références à l'AFP.

## L'AFP : rappels historiques

L'AFP est l'héritière de l'Agence Havas, apparue sous ce nom en 1835. Son histoire – de simple bureau de traduction à agence privée combinant les activités d'agence de presse et d'agence publicitaire, en passant par son statut d'établissement public doté de la personnalité civile et de l'autonomie financière mais de fait sous contrôle direct du gouvernement au lendemain de la Seconde Guerre mondiale – fut longue et mouvementée, et elle a déjà été en grande partie documentée<sup>1</sup>.

L'agence a été créée sous ce nom le 30 septembre 1944 par une ordonnance du gouvernement provisoire signée par le général de Gaulle et les ministres de l'Information et des Finances. Mais elle est alors placée sous le contrôle direct du gouvernement. Les péripéties qui vont lui permettre d'obtenir enfin son indépendance sont nombreuses et c'est le 10 janvier 1957 qu'est promulguée la loi 57-32 portant statut de l'Agence France-Presse. Dans ce nouveau statut, le responsable de l'agence est élu par un conseil d'administration de quinze membres – conseil dans lequel la presse écrite quotidienne détient la majorité par huit sièges (aux côtés de deux représentants de l'audiovisuel public, de trois représentants des services publics usagers et de deux représentants élus par le personnel) – et n'est donc plus nommé par le gouvernement<sup>2</sup>. Si le statut de l'agence a été modifié par la loi du 17 avril 2015, en particulier pour reconnaître ses missions d'intérêt général, il n'en reste pas moins que dans son fonctionnement actuel, l'agence est l'héritière directe de la loi 10 janvier 1957 (**ENCADRÉ**).

<sup>1</sup> Voir en particulier Baron, X. (2014). *Le monde en direct : de Charles-Louis Havas à l'AFP, deux siècles d'histoire*. Cahiers libres. La Découverte, d'où nous nous sommes librement inspirés ici, ainsi que Huteau, J. and Ullmann, B. (1992). *AFP, une histoire de l'Agence France Presse : 1944-1990*, Robert Laffont, Paris.

<sup>2</sup> Notons toutefois que le gouvernement a alors un droit de veto de fait sur la nomination du président de l'AFP puisque un minimum de 12 voix est nécessaire à son élection.

### **Le statut juridique de l'AFP : une exception**

Le régime juridique de l'AFP est *sui generis*. Son statut, promulgué par la loi 57-32 du 10 janvier 1957 portant statut de l'Agence France-Presse, prévoit que l'AFP est un « organisme autonome doté de la personnalité civile et dont le fonctionnement est assuré suivant les règles commerciales ». Cet organisme, qui ne peut être dissous que par une loi, ne compte aucun actionnaire. De plus, l'article 2 prévoit que l'AFP « ne doit, en aucune circonstance, passer sous le contrôle de droit ou de fait d'un groupement idéologique, politique ou économique ». De manière importante, ce statut positionnait l'État comme simple client de l'AFP et garantissait l'indépendance de l'agence.

Le statut de l'AFP a été modifié partiellement en 2012 puis par la loi du 17 avril 2015 afin d'adapter les structures de gouvernance de l'Agence et de clarifier les liens qui unissent celle-ci à l'État. En particulier, cette loi a introduit une obligation de distinction comptable entre les dépenses relevant de missions d'intérêt général de l'Agence et les autres dépenses (voir l'encadré sur la contribution financière de l'État à l'AFP).

Qu'en est-il des deux autres agences de presse mondiales ? Le terme de « *associated press* » a été utilisé pour la première fois aux États-Unis à la fin des années 1840. Entre 1846 et 1848, une demi-douzaine de journaux new-yorkais ont commencé à collaborer pour partager les coûts liés à l'utilisation du télégraphe ainsi que les informations arrivant par bateaux dans la baie de New York<sup>1</sup>. L'objectif de ces journaux était double : d'une part, éviter la duplication des coûts de transmission, de l'autre, obtenir un avantage sur les journaux exclus de l'accord. Ce sont ces derniers qui ont utilisé les premiers le terme de « *associated press* » pour désigner le groupe de journaux alliés. Du XIX<sup>e</sup> siècle à aujourd'hui, l'histoire de AP fut pleine de rebondissements, de

1 Starr, P. (2005). *The Creation of the Media : Political Origins of Modern Communications*, Basic Books.

l'accord de collusion secret passé avec United Press (UP) – une agence de presse concurrente – en 1885 à l'effondrement de l'ancienne AP en 1892 suite à la révélation de cette collusion, en passant par une décision de la Cour suprême en 1945 (*Associated Press v. United States* 326 U.S. 1) déclarant illégal les accords exclusifs entre AP et un nombre réduit de journaux. Aujourd'hui, le statut d'Associated Press (AP) est encore celui d'une coopérative.

Pourquoi l'AFP n'a-t-elle pas, comme son concurrent AP, choisit le statut coopératif ? Cela peut sembler d'autant plus étonnant que l'ordonnance du 30 septembre 1944 du gouvernement provisoire créant l'agence sous son nom actuel précisait qu'« une ordonnance ultérieure fixera les conditions dans lesquelles une agence coopérative d'information sera substituée à l'Agence France-Presse ». La raison en est principalement financière : le statut coopératif – étant donné l'état financier de la presse française – n'aurait sans doute pas permis à l'AFP d'avoir des ressources financières suffisantes. Le soutien financier de l'État est depuis plus de soixante-dix ans indispensable au bon fonctionnement de l'agence.

Initialement créée à Londres en 1851 par Paul Julius Reuter, l'agence Reuters fut de 1930 – date à laquelle Press Association, possédée par la presse locale anglaise, en prit l'intégralité du contrôle – à 1984 – date de la création de Reuters Holdings, compagnie cotée en Bourse – possédée par les journaux eux-mêmes (les journaux locaux furent dès 1941 rejoints par les nationaux) organisés en trust. Aujourd'hui agence de presse canado-américaine, Thomson Reuters (nouveau nom de l'agence depuis le rachat en 2008 de Reuters par Thomson) est une agence purement commerciale<sup>1</sup>.

-----  
Sur l'histoire de Reuters, voir par exemple Mooney, B. et Simpson, B. (2004). *Breaking News : How the Wheels Came off at Reuters*. Wiley.

## Un modèle BtoB

Comment expliquer le rôle prédominant de l'AFP dans la production d'information en France ? Il est intéressant de noter que l'AFP se caractérise par un modèle économique très spécifique et unique dans notre corpus de médias d'information politique et générale. En effet, comme agence de presse, elle repose sur un modèle business-to-business (BtoB). En d'autres termes, l'AFP vend sa production aux autres médias et non directement aux consommateurs, quand les autres médias vendent l'information qu'ils produisent aux consommateurs (modèle BtoC ou business-to-consumer<sup>1</sup>). L'AFP est donc le seul acteur médiatique rémunéré directement pour l'utilisation de ses contenus (même s'il est à noter que le prix auquel ces contenus sont vendus n'est pas fixé par l'AFP comme il le serait par une entreprise commerciale « classique<sup>2</sup> »).

Ainsi, si l'AFP possède un site internet (afp.com), de nombreux fils Twitter et une page Facebook, son objectif n'est nullement d'en maximiser l'audience afin d'augmenter ses revenus publicitaires (l'agence ne perçoit d'ailleurs aucun revenu publicitaire). Sur ces différents supports, l'AFP développe un fil d'actualités mais pointe, dans la majorité des cas, sur des articles de presse ayant fait l'objet d'une reprise de ses dépêches et non sur les dépêches elles-mêmes. L'agence ne se met ainsi pas en concurrence avec ses clients<sup>3</sup>.

1 Vendre est à entendre ici au sens large puisque, comme nous l'avons souligné, la plupart des sites d'information en France donnaient à lire leurs informations gratuitement en 2013 (ils n'avaient pas mis en place de *paypal*). Mais les utilisateurs sont ensuite « vendus » aux annonceurs, ce qui permet aux médias de monétiser leur audience (voir par exemple Angelucci, Ch. et Cagé, J. (2016). « Newspapers in Times of Low Advertising Revenues », *op. cit.*).

2 Les tarifs de l'AFP, depuis la fin de la Seconde Guerre mondiale, ne dépendent en effet plus du volume du service diffusé par l'agence. Selon un principe coopératif, le montant des abonnements est fixé en fonction de l'importance du tirage de chaque journal. L'AFP a toutefois lancé en 2011 un nouveau fil, l'Essentiel-PQR, au volume plus réduit – pouvant être complété par des modules thématiques – mais moins cher. La mise en place de fils « low cost » est d'ailleurs loin d'être spécifique à l'AFP puisqu'elle est pratiquée depuis de nombreuses années par d'autres agences, par exemple AP.

3 Ce qui ne veut pas dire que la présence de l'AFP sur les réseaux sociaux et plus généralement sur Internet ne se soit pas faite sans soulever de nombreux débats, notamment statutaires, car le rôle de l'AFP est historiquement celui d'une agence, fournissant en BtoB des informations aux médias, et non celui d'un média fonctionnant en BtoC, c'est-à-dire en contact direct avec le public.

Les revenus de l'AFP proviennent de l'abonnement des autres médias et de la contribution financière de l'État. L'AFP a intérêt à maximiser le nombre de ses abonnés, en France comme à l'étranger. C'est donc principalement sur la qualité de sa production d'information au niveau tant national qu'international, en termes de couverture, de *breaking news* et de réactivité, que repose son modèle.

*L'AFP, le seul acteur au modèle économique approprié ?*

Ce qui caractérise le modèle économique de l'AFP, c'est de faire payer la reprise de son contenu<sup>1</sup>. Au contraire, les autres médias d'information de notre corpus voient leurs contenus repris sans être payés pour ces reprises.

Cela peut sembler évident, mais la question de la propriété intellectuelle du texte – et en particulier dans le cas de l'information la question de la propriété intellectuelle des faits – est une question extrêmement complexe. Pour les images comme pour la vidéo, les acteurs ont appris à payer pour la réutilisation de contenu. La syndication et le republishing se développent de plus en plus. Mais la plupart de la reprise *littérale* de texte – qui est non négligeable – n'est, elle, pas vraiment réglementée. Ou, pour être plus précis, cette reprise littérale pourrait être rattachée à un certain nombre de textes de loi sur la propriété intellectuelle mais *de facto* pratiquement aucune poursuite juridique n'a été menée sur ce terrain entre médias d'information<sup>2</sup>.

Or, être payé pour la reprise de son contenu, cela revient à monétiser ce dernier. Autrement dit, pour un média, à obtenir des incitations financières à investir dans la production (coûteuse) d'une information originale. Incitations qui se sont réduites cruellement au cours des dernières années avec, d'une part, l'effondrement des recettes publicitaires et, d'autre part, une tendance croissante des citoyens à ne pas vouloir payer pour de l'information généraliste en ligne<sup>3</sup>.

1 Le paiement ne se faisant pas « à la pièce » mais à travers les abonnements.

2 Les batailles juridiques ont été plus nombreuses entre médias d'information générale d'une part et agrégateurs – comme Google News – d'autre part, mais la problématique dans ce cas est différente.

3 D'après les estimations du dernier rapport du Reuters Institute (*Reuters Institute Digital News Report 2016*), seuls 11% des Français sont prêts en 2016 à payer pour de l'information en ligne.

*Le développement du copié-collé... à travers le filtre de l'AFP?*

Ce que nous a montré l'exemple du *Monde* et des écoutes de la NSA que nous avons détaillé plus haut, c'est que le copié-collé se fait souvent à travers le filtre de l'AFP, ce qui revient d'une certaine manière à le « légaliser ». *Le Monde* sort une information exclusive : l'AFP est en droit de la reprendre et de la diffuser, ce qui est d'ailleurs dans l'intérêt du *Monde* qui avait mis cette information sous embargo en amont avec l'AFP. Comme les dépêches de l'AFP « ressemblent » très fortement à l'article original du *Monde*<sup>1</sup>, quand les autres médias reprennent à leur tour presque littéralement sur leur site les dépêches de l'AFP – ce qu'il est légitime pour eux de faire à partir du moment où ils sont abonnés au fil AFP –, les articles qu'ils publient ressemblent pour l'essentiel à l'article original du *Monde*.

Cela n'est pas problématique tant que *Le Monde* est cité comme la source originale de l'information. Or cela est loin d'être toujours le cas.

## Une mission d'intérêt général

L'AFP se caractérise non seulement par son modèle économique particulier – en BtoB – mais également par sa mission d'intérêt général. Cette mission d'intérêt général a été reconnue par la Commission européenne en 2015 (**ENCADRÉ**). Depuis 2015, la contribution financière de l'État en faveur de l'AFP se décompose en un abonnement aux services de l'AFP d'une part et en une subvention pour compensation du coût net des missions d'intérêt général d'autre part. Le financement de l'État en faveur de l'AFP au titre des missions d'intérêt général a représenté près de 35% des recettes de l'agence en 2015.

---

Ce chiffre est de 9% aux États-Unis et de seulement 7% au Royaume-Uni.

1 Une manière positive de le voir serait de dire qu'il n'y pas une infinité de façons de formuler un même fait.

### **La contribution financière de l'État à l'AFP : concurrence déloyale ou mission d'intérêt général ?**

Conformément à la loi du 10 janvier 1957 portant statut de l'AFP, une convention avait fixé en 1958 les modalités d'abonnement de l'État – simple client de l'AFP – aux services de l'AFP, en référence aux tarifs pratiqués alors par l'AFP auprès de ses clients de la presse quotidienne. (La convention prévoyait que « en rémunération de l'ensemble des services [...], l'État versera à l'Agence France Presse le montant de 383 abonnements d'après le tarif qui sera appliqué aux abonnements souscrits par les entreprises de presse française exploitant un journal quotidien ayant un tirage de 180 000 exemplaires ».)

En 1958, l'abonnement annuel payé par l'État représentait 59% du chiffre d'affaires de l'AFP. Il ne cessera par la suite de décroître, passant en dessous des 50% en 1991, pour s'établir à 41% en 2013 et à environ un tiers aujourd'hui.

En 2010, une plainte contre la France a été déposée auprès de la Commission européenne par une agence d'information allemande (DAPD) pour aide d'État présumée en faveur de l'AFP. Son instruction par la Commission européenne a abouti à la conclusion que, au regard de leur montant, les abonnements de l'État français aux services de l'AFP constituent bien une aide d'État existante. Sous certaines conditions agréées en mars 2014 par la Commission européenne et la France, cette aide d'État est néanmoins compatible avec le droit de l'Union européenne.

Ainsi, depuis 2015, du fait de la mise en conformité du financement de l'AFP avec les règles européennes de concurrence, la contribution financière de l'État en faveur de l'AFP se décompose en un abonnement commercial aux services de l'AFP (abonnement dont le montant doit être conforme aux grilles tarifaires de l'AFP) et en une subvention pour compensation du coût net des missions d'intérêt général, clarifiant ainsi les activités de l'AFP relevant d'une mission d'intérêt général et ses activités de nature commerciale.

La loi du 10 janvier 1957, telle qu'elle résulte des modifications issues de la loi du 17 avril 2015, prévoit à son article 12 que la Commission financière s'assurera annuellement que les compensations versées par l'État n'excèdent pas les coûts nets générés par l'accomplissement des missions d'intérêt général. La question des contours exacts de ces missions d'intérêt général et de son évaluation financière est maintenant officiellement posée; elle n'est sans doute pas près d'être définitivement réglée.

Les missions d'intérêt général sont définies aux articles 1<sup>er</sup> et 2 de la loi du 10 janvier 1957, telle que modifiée par la loi du 22 mars 2012. En particulier, l'AFP est soumise aux trois obligations fondamentales suivantes :

1. Elle ne peut en aucune circonstance tenir compte d'influences ou de considérations de nature à compromettre l'exactitude ou l'objectivité de l'information; elle ne doit, en aucune circonstance, passer sous le contrôle de droit ou de fait d'un groupement idéologique, politique ou économique.
2. Elle doit, dans toute la mesure de ses ressources, développer son action et parfaire son organisation en vue de donner aux usagers français et étrangers, de façon régulière et sans interruption, une information exacte, impartiale et digne de confiance.
3. Elle doit, dans toute la mesure de ses ressources, assurer l'existence d'un réseau d'établissements lui conférant le caractère d'un organisme d'information à rayonnement mondial.

La reconnaissance des missions d'intérêt général de l'AFP – dont les coûts nets sont donc compensés par une subvention de l'État – est particulièrement intéressante dans le cadre de la réflexion que nous menons ici sur la production d'information et les moyens de préserver les incitations des médias.

## CHAPITRE 5

### LA MONÉTISATION DE L'INFORMATION

Pourquoi s'intéresser à la monétisation de l'information ? Non pas parce que nous pensons que les médias sont des entreprises comme les autres dont l'objectif principal devrait être de maximiser leur rentabilité – bien au contraire. Mais parce que ce sujet nous concerne tous comme citoyens désireux de nous informer et de participer ainsi au bon fonctionnement de la démocratie. L'information est certes un bien public, mais c'est un bien public qui coûte cher à produire : il faut réussir à la monétiser pour pouvoir continuer à le faire.

Or quelles sont les solutions qui s'offrent à une société prise dans son ensemble pour monétiser l'information ? Elles sont en nombre limité :

1. Les utilisateurs directement à travers leur acte d'achat ou d'abonnement.
2. Les publicitaires (mais qui se préoccupent du nombre d'utilisateurs présents sur les différentes plateformes).
3. L'État (c'est-à-dire indirectement les utilisateurs / citoyens).
4. Des actionnaires prêts à perdre de l'argent et dont on peut interroger les motivations.

Ce à quoi l'on peut rajouter les autres médias à travers la syndication de contenu, mais cela nous ramène indirectement aux utilisateurs<sup>1</sup>.

---

<sup>1</sup> Tout comme d'ailleurs se rattachent aux consommateurs/citoyens et/ou aux publicitaires les nouvelles formes de diversification des sources de revenu des médias, de l'organisation de conférences à la pratique de la « marque blanche ».

On ne s'arrêtera pas ici sur le cas des actionnaires qui a fait l'objet d'un précédent ouvrage par l'un des auteurs de ce livre<sup>1</sup>. Concentrons-nous pour commencer sur les utilisateurs. Même si en France en 2013 la plupart des médias d'information offraient la grande majorité de leur contenu en ligne gratuitement, le nombre d'utilisateurs est en effet une variable clef pour les publicitaires.

## À qui profite la production d'information originale ?

Comment l'audience des différents sites réagit-elle à leur production d'information ? Nous avons mis en regard, jour par jour, le nombre de visiteurs uniques et la production d'information quotidienne des différents médias<sup>2</sup>.

### *Les variations de l'audience agrégée*

Mais intéressons-nous tout d'abord à la façon dont l'audience agrégée réagit à l'actualité.

L'audience agrégée est fonction du jour de la semaine<sup>3</sup> ainsi que de la période de l'année. On observe une baisse d'audience très nette durant les mois d'été et autour de Noël. Cela est vrai non seulement de l'audience des sites d'information mais plus largement de l'audience agrégée sur Internet (**FIGURE 20**).

L'audience varie également en fonction du nombre et de l'importance des événements de la journée. Comment mesurer l'importance d'un événement ? La « pression médiatique » (ou « *news pressure* ») d'un événement sur un jour donné est définie comme la part de l'ensemble des articles publiés consacrés à cet événement<sup>4</sup>. Il suffit par

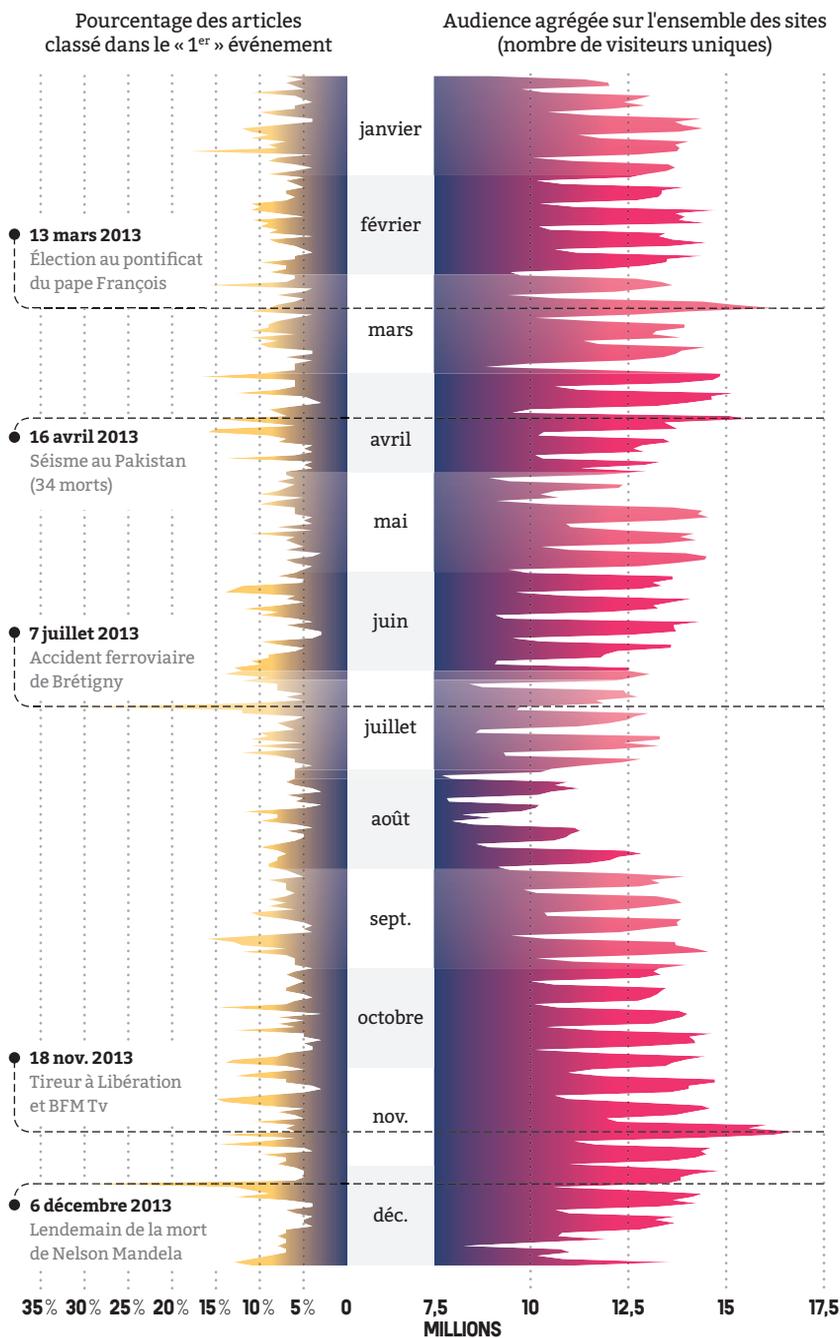
1 Cagé, Julia (2015) : *Sauver les médias*, *op. cit.*

2 Ces données proviennent de l'OJD (aujourd'hui appelée ACPM – Alliance pour les Chiffres de la Presse et des Médias – à la suite de sa fusion avec AudiPresse). L'OJD est une organisation en charge de certifier historiquement la diffusion des journaux et aujourd'hui également les audiences de leur site internet.

3 Pablo Boczkowski a ainsi montré que l'actualité en ligne était consultée principalement au bureau, c'est-à-dire sur le temps de travail. Boczkowski, P. J. (2010). *News at Work*, *op. cit.*

4 Voir par exemple Eisensee, Th. et Strömberg, D. (2007). « News Droughts, News Floods, and U. S. Disaster Relief », *The Quarterly Journal of Economics*, 122 (2), 693-728 ; Durante, R. et Zhuravskaya, E. (2015). « Attack When the World Is Not Watching? International Media and the Israeli-Palestinian Conflict », CEPR Discussion Papers 10750 ; Philippe, A. and Ouss, A. (2015). « “No hatred or Malice, fear or affection” : Media and Sentencing » , Document de travail.

→ **Figure 20** **PRESSION MÉDIATIQUE ET AUDIENCE**



Source des données : Alliance pour les Chiffres de la Presse et des Médias (ACPM).

exemple de penser en 2013 à la pression médiatique qui a accompagné la mort de Nelson Mandela ou encore l'accident ferroviaire en gare de Brétigny. Cela apparaît très fortement visuellement sur la **FIGURE 20** qui représente chaque jour la part des articles consacrés au « principal » événement de la journée.

Il est intéressant de noter cependant que les pics d'audience sont loin de correspondre parfaitement à la place accordée par les médias à certains événements. Ainsi, alors que la pression médiatique associée à l'accident de Brétigny est extrêmement forte le 7 juillet, l'audience des sites internet est, elle, relativement faible (ce qui n'est pas très étonnant s'agissant d'un dimanche de juillet). Au contraire, l'élection du Pape François le 13 mars 2013 semble générer davantage l'intérêt des internautes (si l'on en croit l'audience agrégée) que l'attention de médias en ligne.

L'analyse économétrique dont nous allons ici simplement présenter les principaux résultats<sup>1</sup> étudie les variations d'audience agrégée en fonction de quatre déterminants<sup>2</sup> : (1) le nombre d'événements ; (2) le pourcentage des articles classés dans le « premier » événement ; (3) le pourcentage des articles classés dans le « deuxième » événement ; et (4) le pourcentage des articles classés dans le « troisième » événement (le rang des événements étant défini ici en fonction du nombre d'articles et du nombre de médias traitant de chaque événement).

L'audience agrégée augmente de façon statistiquement significative avec le nombre d'événements et la pression médiatique associée au premier événement. Par exemple, une augmentation de 1% du nombre d'événements couverts conduit à une augmentation d'environ 0,15% du nombre total de visiteurs. Dit autrement, un événement supplémentaire accroît le nombre de visiteurs *totaux* (sur l'ensemble des sites d'information de notre corpus) en moyenne de 20 000 à 42 000 en fonction des spécifications. De plus, une augmentation d'un point de pourcentage de la part des articles consacrés au principal événement de la journée augmente l'audience agrégée de 45 000 à 71 000 visiteurs.

L'audience agrégée varie donc avec l'actualité de chaque journée (ce qui – il faut bien le reconnaître – n'est pas très étonnant). Mais

1 Comme précédemment, le lecteur intéressé par les détails économétriques pourra se reporter à l'article technique en anglais.

2 En contrôlant pour le jour de la semaine et le mois de l'année.

dans quelle mesure cette augmentation de l'audience agrégée bénéficie-t-elle particulièrement aux médias qui contribuent le plus à la production d'information originale en rapport avec cette actualité ? En d'autres termes, aux médias qui supportent le coût de la production de cette actualité ?

*Les déterminants de la part d'audience de chaque média d'information*

Nous avons calculé pour chacun des médias sa part d'audience quotidienne. En moyenne, le nombre de visiteurs quotidiens varie très fortement d'un média à l'autre, de moins de 20 000 visiteurs quotidiens pour les sites internet des petits titres de presse quotidienne régionale comme *La République du Centre* à plus d'un million pour des journaux nationaux comme *Le Monde* ou *Le Figaro*. Si l'on s'intéresse tout d'abord à la corrélation entre le nombre moyen de visiteurs uniques des sites et le nombre d'événements pour lesquels ils ont été *news breakers* en 2013, on s'aperçoit que celle-ci est très forte (**FIGURE 21**).

Mais ce qui nous intéresse, ce n'est pas tant d'expliquer la différence d'audience entre *La République du Centre* et *Le Figaro*, que les variations quotidiennes d'audience de ces médias. Car de nombreux facteurs peuvent venir expliquer cette corrélation, en particulier le fait que les médias à l'audience la plus élevée reçoivent en moyenne davantage de revenus publicitaires, ce qui leur permet d'investir plus dans la production d'information et de produire plus de scoops.

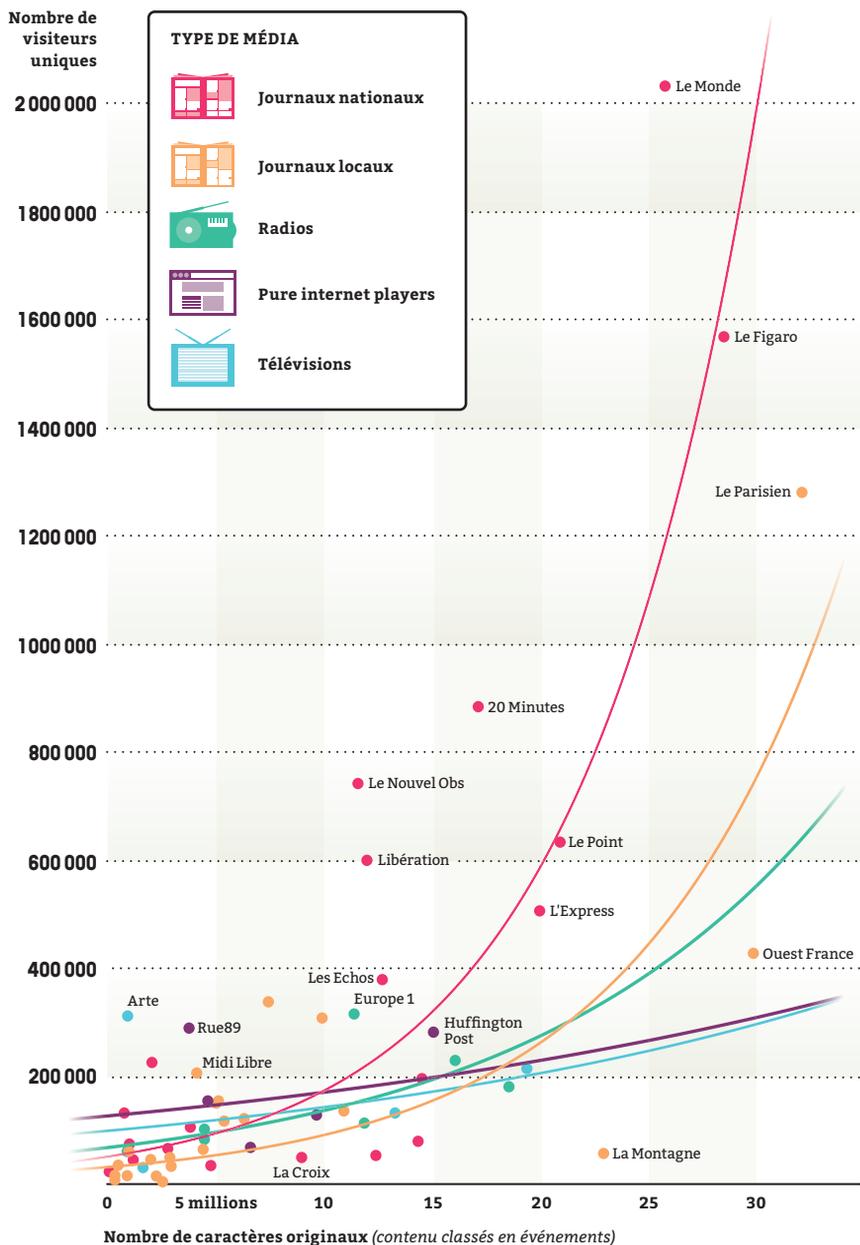
Pour déterminer l'impact *causal* de la production d'information originale sur l'audience des sites internet des médias, nous avons régressé la part d'audience quotidienne de l'ensemble des médias de notre corpus sur un certain nombre de déterminants que nous allons détailler. Une régression, en économétrie, permet d'estimer les paramètres de la fonction qui relie les variables explicatives à une variable expliquée.

Notons tout d'abord que nous avons inclus dans ces régressions économétriques non seulement des « effets fixes jour » mais également des « effets fixes média ». Au-delà de la terminologie qui manque cruellement de poésie, cela signifie que les effets que nous avons mis à jour peuvent être considérés comme des effets causaux : l'augmentation de telle dimension de la production originale d'information augmente *causalement* la part d'audience du média.

→ Figure 21

## LA CRÉATION DE CONTENU ORIGINAL ATTIRE-T-ELLE DES VISITEURS ?

Nombre de visiteurs uniques et nombre de caractères originaux dans les contenus classés



Les déterminants que nous avons introduits dans notre analyse sont les suivants : (1) le contenu total non classé dans des événements<sup>1</sup> ; (2) le contenu non original classé dans des événements ; (3) le contenu original classé dans des événements ; et (4) le nombre d'événements pour lesquels le média est le *news breaker*.

Les résultats que nous présentons ici sont surprenants. Le seul déterminant causal statistiquement significatif de la part d'audience des médias est la production de contenu original. Le nombre de *breaking news* n'affecte pas l'audience en ligne des médias. On peut penser que la production de contenu copié-collé rebute les lecteurs, même si l'effet mesuré est extrêmement faible et peu significatif. Des limites du maniement des ciseaux ?

Cependant, s'il est statistiquement significatif, l'effet de la production originale d'information est néanmoins quantitativement faible : en moyenne quotidienne, une augmentation de 1% de la production de contenu original n'augmente que de 0,018% le nombre de visiteurs uniques d'un site. Autrement dit, les utilisateurs ne « récompensent » presque pas les producteurs d'information originale. L'information en ligne étant substituable – tous les sites reprenant simultanément les mêmes informations souvent avec les mêmes mots – celui qui la produit ne parvient donc plus à la monétiser, tout du moins en termes d'audience.

Les utilisateurs ne réagissent pas plus au nombre de *breaking news* mises en ligne par un média dans la journée. Toutefois, si l'on considère cette variable de manière binaire – le fait de ne publier aucune *breaking news* versus au moins une – alors on s'aperçoit que le fait pour un média de publier au moins une *breaking news* au cours de la journée lui permet d'augmenter son nombre de visiteurs uniques du jour de 1,4%. Autrement dit, même si le coût de court terme est plus élevé – et si le bénéfice reste quantitativement faible – les éditeurs devraient investir un peu plus dans l'investigation et un peu moins dans les ciseaux.

Ces résultats ont deux implications majeures. D'une part, l'importance de la réputation pour un média. D'autre part, le fait que la baisse notable du nombre de journalistes, et en particulier la diminution de la taille moyenne des rédactions, est au moins en partie due aux innovations technologiques qui ont accompagné la révolu-

<sup>1</sup> Nous avons choisi de faire cette analyse en nombre de caractères plutôt qu'en nombre d'articles du fait de la taille très variable des articles.

tion numérique, et notamment au fait qu'il est plus que jamais facile de « voler » à la fois du contenu et des visiteurs à ses concurrents. Ce qui peut, à première vue, apparaître comme une bonne nouvelle pour les citoyens consommateurs – il leur est aujourd'hui possible d'accéder gratuitement d'un clic et en seulement quelques secondes à l'ensemble des informations disponibles, peu importe le site internet sur lequel ils se trouvent – pourrait malheureusement à terme se retourner contre eux (et contre la démocratie). Car sans incitations monétaires suffisantes pour les médias à produire de l'information originale, celle-ci pourrait dans le futur venir à se raréfier. Ainsi, les médias réalisant qu'en termes d'audience ils ne bénéficient pas de leur investissement dans la production d'information originale, pourraient choisir d'en produire moins, et de diminuer en conséquence la taille de leurs rédactions.

Les données que nous avons utilisées ne couvrent qu'une seule année et ne nous permettent pas d'analyser dans quelle mesure une telle raréfaction de l'information originale s'est produite dans les années récentes ou serait en train de poindre à l'horizon. Mais la baisse observée du nombre de journalistes couplée à la corrélation entre nombre de journalistes et production d'information originale le laisse malheureusement à penser. Confirmant ainsi par exemple les résultats de James Hamilton qui, dans le cadre de son étude de l'investigation journalistique aux États-Unis, a montré que le nombre de soumissions d'articles au prix de l'Investigative Reporters & Editors (IRE) par les journaux locaux américains s'est effondré en même temps que le nombre de leurs journalistes<sup>1</sup>.

*Les utilisateurs récompensent-ils les médias pour leur réactivité ?*

Au-delà de la seule dimension du *breaking news*, il est intéressant de se demander dans quelle mesure les utilisateurs récompensent les médias pour leur réactivité à couvrir une actualité. Et l'on pourrait même se demander dans quelle mesure ils pénalisent les médias trop lents à en traiter.

Pour évaluer l'importance de ce phénomène, nous sommes rentrés dans le détail de chaque événement et nous avons calculé le temps

<sup>1</sup> Hamilton, J. T. (2016). *Democracy's Detectives : The Economics of Investigative Journalism*, op. cit.

mis par chaque média pour couvrir une actualité (quand il la couvre). Nous avons ensuite étudié dans quelle mesure la part d'audience d'un média augmentait avec sa réactivité<sup>1</sup>.

Plus la réactivité d'un média est faible pour un événement donné, plus sa part d'audience l'est également, et ce quel que soit l'indice de réactivité utilisé (le rang du média dans l'ordre de publication ou sa réactivité calculée en secondes). Ce résultat renforce l'analyse de Boczkowski (2010) qui soulignait déjà la préoccupation des journalistes à réagir « dans la minute » à un événement de peur de voir leur audience s'écrouler immédiatement et irrémédiablement<sup>2</sup>. Il rationalise également la stratégie éditoriale de nombreux médias qui consiste à couvrir au plus vite – et donc en recourant au copié-collé – chaque événement, même si cela aboutit à l'arrivée à une homogénéisation des contenus publiés sur Internet.

Cependant, si cet effet est statistiquement significatif, sa magnitude est néanmoins très faible. De plus, nos résultats sur l'impact causal de la production d'information originale sur la part d'audience des médias montrent que l'audience décroît de manière inversement proportionnelle au taux de copie. La tentation du gain de temps de réaction par l'intermédiaire du copié-collé peut donc s'avérer contre-productive, au final, en termes d'audience.

*Des effets non seulement faibles mais également non persistants*

Par ailleurs, si les médias profitent d'une (faible) augmentation de leur audience lorsqu'ils innovent, et encore plus lorsqu'ils innovent vite, cet effet est de courte durée.

Pour le mesurer, nous avons étudié l'impact de la production d'information originale et de la réactivité d'un média un jour donné sur son audience du lendemain (en contrôlant bien sûr pour sa performance du jour). Or il apparaît clairement que les utilisateurs ont

1 Avec une limite cependant : la réactivité est calculée ici au niveau de l'événement, mais l'audience des sites n'est disponible qu'au niveau de la journée; autrement dit, nous ne connaissons pas pour chaque article publié à l'intérieur d'un événement le nombre de lecteurs de l'article et ne pouvons donc qu'étudier l'effet de la réactivité *moyenne* d'un média au cours d'une journée sur sa part d'audience

2 Cela est également consistant avec les résultats obtenus par Ignacio Franceschelli qui a étudié la production d'information et l'audience hors ligne et en ligne des deux principaux journaux argentins. Franceschelli, I. (2011). « When the Ink is Gone : The Transition from Print to Online Editions. » Technical report, Northwestern University.

la mémoire courte : la production d'information originale de la veille (pas plus que les autres dimensions du travail du média) n'affecte en rien la part d'audience du lendemain.

### **Comment expliquer la faiblesse des variations d'audience ?**

Alors que l'audience agrégée réagit fortement à la pression médiatique, les parts d'audience des différents médias sont relativement stables en moyenne, peu importe la production d'information un jour donné. Comment expliquer cette stabilité ? Plusieurs explications, qui peuvent sembler opposées à première vue, peuvent permettre de la rationaliser. D'une part, la combinaison de la rapidité de la propagation de l'information et de l'utilisation des réseaux sociaux et des agrégateurs comme porte d'entrée vers l'information fait que les utilisateurs ne sont plus incités à se rendre en priorité sur le site des *news breaker*. D'autre part, les utilisateurs ont leurs « marques » préférées : peu importe qui produit originellement l'information, ils se rendront sur le site de ces médias. Effet de marque ou volatilité : dans un cas comme dans l'autre, le *news breaker* n'est pas récompensé.

*Quelle(s) porte(s) d'entrée pour la consommation d'information ?*

En moyenne, 46% de l'audience en ligne provient des moteurs de recherche, comme Google, à travers l'utilisation de mots-clefs<sup>1</sup>. Seule 27% de l'audience parvient directement sur le site internet des médias. Une part importante des visiteurs (près de 10%) se rend également sur les sites à partir des réseaux sociaux, comme Facebook, Twitter ou encore Reddit.

Autrement dit, les utilisateurs viennent majoritairement des réseaux sociaux et autres agrégateurs. Ils tendent donc à cliquer sur

---

<sup>1</sup> Ces données proviennent de SimilarWeb, une entreprise de technologie de l'information qui commercialise des données de consommation d'internet. Les chercheurs que nous sommes ne peuvent d'ailleurs que déplorer le fait de devoir dépenser de l'argent de recherche pour acheter – à des seules fins de recherche – des statistiques sur la consommation des sites internet des médias d'information générale français à une entreprise étrangère. Fournir de telles statistiques aux chercheurs – avec tous les engagements de confidentialité nécessaires de la part de ceux-ci – pourrait être une contrepartie demandée aux médias en échange des aides à la presse qu'ils reçoivent.

les liens les plus visibles (ceux que Google News leur proposera par exemple en premier) sans se demander qui est à la source de l'information.

Certes, la réactivité est importante pour les agrégateurs – l'algorithme de Google favorise ainsi les médias qui sont les premiers à couvrir un sujet –, mais ce qui l'est plus encore, c'est que l'information qui nous est donnée à consommer est passée au « filtre » de ces derniers, comme l'algorithme Edge Rank de Facebook. Chaque individu est ainsi enfermé dans une bulle informationnelle qui lui est propre et dont il n'a le plus souvent pas conscience, et c'est donc en méconnaissance de fait qu'il choisira de ne pas récompenser les médias les plus « méritants ». L'élection de Donald Trump a notamment relancé le débat sur le rôle et la pertinence de ces algorithmes de filtrage automatique des articles, notamment sur Facebook. Outre la bulle informationnelle ainsi créée, la hiérarchisation et la vérification des informations sont également fortement critiquées. Or pour ces tâches, on est encore loin d'avoir trouvé un meilleur algorithme que le journaliste<sup>1</sup> !

### **Les recettes publicitaires en ligne, une source de revenu pour les médias d'information ?**

Même si la production d'information originale conduisait à des variations d'audience plus importantes que celles que nous observons pour 2013, cela n'impliquerait pas pour autant que les médias soient en capacité de monétiser cette audience, en tous cas en l'absence de mur payant.

Les revenus publicitaires ne représentaient en effet en 2013 que 5% des revenus des sites en ligne. Pourquoi ? Non parce que le marché publicitaire en ligne serait inexistant. Au contraire, Internet est devenu aux États-Unis – et deviendra sans doute en France dans un futur proche – le premier support pour les dépenses publicitaires. Mais parce que la plus grande partie de ce marché est aujourd'hui capturée par les quatre principaux acteurs du web (et en particulier Google

<sup>1</sup> Nous renvoyons les lecteurs intéressés par ce sujet aux travaux passionnants de Dominique Cardon et de Eli Pariser : Cardon, D. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Seuil, « La République des idées ». Pariser, E. (2011). *The Filter Bubble : How the New Personalized Web Is Changing What We Read and How We Think*, Penguin Publishing Group.

et Facebook) et échappe donc aux médias traditionnels comme aux *pure players*. De plus, tout laisse à penser que si rien n'est fait, ce phénomène n'est pas près de s'inverser, bien au contraire.

Il faut également ajouter à cette capture par un nombre réduit d'acteurs de la majorité des dépenses publicitaires en ligne, l'utilisation croissante des ad-blockers – qu'il s'agisse d'Adblock Plus ou de Brave, parmi d'autres – par les consommateurs d'information. D'après les estimations du dernier rapport du Reuters Institute<sup>1</sup>, 30% des internautes utilisent aujourd'hui un ad-blocker en France, chiffre qui atteint 38% en Pologne ou encore 36% en Grèce.

Une véritable guerre se joue ainsi aujourd'hui entre les ad-blockers d'une part, qui affirment redonner aux utilisateurs leur liberté<sup>2</sup>, et les éditeurs, les agrégateurs et les réseaux sociaux de l'autre, qui se retrouvent dans cet affrontement du même côté du champ de bataille. Août 2016 a ainsi vu un affrontement ouvert entre Facebook et Adblock Plus, Facebook ayant dans un premier temps mis en place un système publicitaire rendant inefficace l'ad-blocker avant que ce dernier réussisse en moins de deux jours à le contourner.

Qui sortira gagnant de cette bataille technologique ? La question reste entièrement ouverte. Mais ce qui se joue ici, c'est bien la question cruciale du financement de l'information. Adblock Plus a ainsi lancé au printemps 2016 en association avec Flatlr (un dispositif ancien mais relativement peu utilisé de donation par micropaiements) un outil de micropaiement visant à inciter les internautes à payer pour le contenu qu'ils lisent sur le site des médias. Nous aurons l'occasion de revenir sur cette question du financement beaucoup plus longuement dans le chapitre 7 consacré aux solutions.

1 Reuters Institute Digital News Report 2016.

2 Le discours libertaire a cependant ici ses limites, et ce sont celles de la réalité des bloqueurs de publicité qui sont pour la plupart des entreprises privées à vocation lucrative. Brave ne cherche ainsi pas uniquement à supprimer les bannières publicitaires sur les écrans des internautes... mais à les remplacer par des publicités qu'il aurait lui-même vendues.



## CHAPITRE 6

### L'IMPORTANCE DES CITATIONS ET DE LA RÉPUTATION

Si les médias ne tirent aucun bénéfice (direct) en termes d'audience de leur production d'information originale en ligne, alors pourquoi certains d'entre eux continuent-ils à en produire et ne font-ils pas plutôt le choix de se comporter comme des « passagers clandestins », en se contentant de reproduire sur leur site l'information produite par d'autres ?

On peut bien sûr faire l'hypothèse – un peu naïve – que tous les patrons de médias seraient des humanistes ayant pour seul objectif de mettre à jour la vérité et d'informer au mieux les citoyens, quel qu'en soit le coût et quitte à en être pour leurs frais. Mais une telle hypothèse étant loin d'être satisfaisante, il nous semble nécessaire de chercher la réponse ailleurs. Et elle se trouve peut-être du côté de la réputation.

#### Références et réputation

La production d'information originale peut en effet être une approche stratégique pour un média, qui tirera de cette production la qualification de « média de référence » ou de média à fort « capital symbolique » pour reprendre la terminologie utilisée par Patrick Champagne dans son dernier ouvrage<sup>1</sup>. Matthew Genztkow et Jesse Shapiro ont ainsi montré que si la concurrence sur le marché des

---

1 Champagne, P. (2016). *La double dépendance. Sur le journalisme*, op. cit.



médias permet d'« atteindre la vérité », c'est en grande partie du fait de la recherche par les médias de gains de réputation<sup>1</sup>.

Toutefois, pour que les médias puissent tirer un bénéfice en termes de réputation de la production d'une information originale, encore faut-il que les médias concurrents respectent les procédures de citations et de crédits<sup>2</sup>. Car si n'est pas rendu à César ce qui lui appartient, l'investissement risque d'être en pure perte. Dans quelle mesure est-ce le cas ?

### *Un algorithme de détection de citations*

Pour étudier cette question, nous avons développé un algorithme de détection de références ou citations qui nous permet, dans chaque article, de détecter les références à des médias comme source de l'information (**ENCADRÉ**).

#### **L'algorithme de détection de citations**

Nous appelons citation le fait pour un média de créditer un autre média comme étant la source d'une information traitée dans un article. Il s'agit de tournures de phrases telles que : « comme révélé par *Le Monde* », « selon un journaliste de l'AFP » ou encore « le site *Mediapart* avait publié le 28 avril un document ». Il faut en particulier être capable de distinguer les cas où le média est référencé en tant que source d'une information des cas où il est partie prenante de l'événement médiatique. Ainsi, pour faire écho à l'actualité au moment où nous écrivons ces lignes, notre algorithme doit faire en sorte de ne pas attribuer de citation à l-Télé dans les nombreux articles qui couvrent la grève qui oppose la rédaction de la chaîne d'information en continu à son actionnaire.

L'algorithme est décomposé en deux grandes étapes :

1 Gentzkow, M. et Shapiro, J. M. (2008). « Competition and Truth in the Market for News », *Journal of Economic Perspectives*, 22 (2), 133-154.

2 De plus, juridiquement, l'« exception de courte citation » en France est soumise à la mention de la source (ainsi que du nom de l'auteur).

1. Nous repérons toutes les apparitions des noms de médias dans les articles à l'aide d'un algorithme de détection d'entité nommée (là encore, il s'agit d'un composant classique en TAL, appelé Named Entity Recognition). En plus des médias de notre corpus, nous avons également inclus environ 80 médias supplémentaires dans cette étape dont une bonne part de médias étrangers. Chaque média peut être reconnu par son nom, son site web ou une périphrase (« le quotidien du soir » ou « l'hebdomadaire satirique » sont ainsi assez fréquents).

2. Pour chaque détection du nom d'un média, nous déterminons à l'aide d'un ensemble de règles permettant de détecter certaines formes linguistiques, si la phrase est bien une citation selon notre définition. Le travail d'écriture de ces règles est relativement long et est fait manuellement à partir de documents issus de l'étape précédente. Nous avons ainsi isolé une soixantaine de verbes impliqués dans ce type de citations (affirmer, dire, écrire, dévoiler, mentionner...) et les contextes dans lesquels ils sont utilisés. Une fois ces règles codées dans un logiciel ad hoc, nous pouvons retrouver les citations dans l'ensemble de notre corpus.

Nous avons obtenu un total de 461 336 citations en 2013, 48% d'entre elles étant des références à l'AFP – la surreprésentation de l'AFP s'expliquant en partie par le nombre élevé d'articles écrits « avec AFP » – et 11% à des médias étrangers (comme Associated Press, Reuters, le *Guardian* ou encore le *New York Times*).

Les médias étrangers sont en effet parfois la source de l'information et quand les médias français couvrent l'événement, ils les mentionnent comme telle. C'est ainsi l'agence Reuters qui a annoncé le 20 juin 2013 le décès de l'acteur américain James Gandolfini, star des *Soprano*; quand *Capital* écrit un article sur le sujet, le mensuel le mentionne clairement : « L'acteur est décédé d'une probable crise cardiaque, a déclaré à Reuters une porte-parole de la chaîne » (la chaîne étant ici la chaîne américaine HBO qui diffusait les *Soprano*). Les médias étrangers tendent en particulier – ce qui n'est pas sur-

prenant – à avoir la primeur des événements se produisant dans leur pays. Ainsi, c'est le *New York Times* qui s'est procuré et a dévoilé le 18 septembre 2013 un rapport de police sur « le tireur de Washington », auteur d'une fusillade dans la capitale américaine le 16 septembre. L'AFP, qui publie une première dépêche sur le sujet à 20h27, se contente ainsi de reprendre l'information révélée par le quotidien américain et note que le tueur « entendait des voix », selon un rapport de police *dévoilé par le New York Times* ».

Nous avons également mis à jour un nombre non négligeable de références à des médias ne faisant pas partie de notre base de données car n'ayant pas de site internet, en particulier le *Canard Enchaîné*<sup>1</sup>. Le *Canard Enchaîné* doit une grande partie de son succès à sa réputation ; jusqu'à la création de *Mediapart*, il était l'un des principaux médias en France à sortir des « affaires<sup>2</sup> ». Le *Canard* compte aujourd'hui 36 journalistes – un chiffre relativement stable depuis le milieu des années 1980 – qualifiés et bien rémunérés. Ce nombre peut sembler faible, mais le journal ne compte que 8 pages – en noir et blanc ! – et ne traite pas directement de l'actualité chaude. Pour ne prendre qu'un exemple, c'est le *Canard* qui a révélé dans son édition du mercredi 6 février 2013 que Bernard Arnault – « merci patron ! » – avait divisé par sept sa taxe sur la succession. Par construction, l'article de l'hebdomadaire satirique n'apparaît pas dans notre base de données. Mais quand l'AFP, la première à couvrir cet événement, fait une dépêche sur le sujet à 18h24 (« Bernard Arnault a réglé sa succession avec des droits réduits drastiquement (presse) ») elle mentionne que l'information a été révélée par le *Canard Enchaîné* : « a affirmé mercredi *Le Canard Enchaîné* ». De même, *Le Parisien*, *Le Journal du Dimanche*, France 24, Europe 1, *Nice Matin* ou encore *Challenges*, qui reprennent tous l'information dans l'heure, font référence au *Canard Enchaîné* comme source de l'information.

Au final, sur les un peu plus de 850 000 articles classés dans des événements pour l'année 2013, 42,5% contiennent au moins une référence à un autre média comme source l'information. Les articles non classés dans des événements contiennent relativement moins

1 Le site internet du *Canard Enchaîné* est de création récente et ne donne à voir que la une de son édition papier. De même que son tout nouveau compte Twitter.

2 Il est d'ailleurs intéressant de noter que la réputation du *Canard Enchaîné* l'aide aussi dans sa production d'information originale, puisque beaucoup d'affaires lui sont aujourd'hui « amenées » (ce qui n'enlève rien au travail des journalistes qui doivent par la suite vérifier la pertinence des révélations et créer l'événement médiatique).

de citations (seulement 14% d'entre eux en contiennent une), ce qui n'est pas étonnant puisque le contenu de ces articles est relativement « indépendant<sup>1</sup> ».

*Les médias les plus copiés reçoivent-ils davantage de citations ?*

En moyenne, chaque média inclus dans notre base de données est cité un peu plus de 4 500 fois en 2013<sup>2</sup>, 1 900 fois si l'on ne considère par l'AFP. Les radios et les télévisions arrivent en tête des médias les plus cités (plus de 12 700 citations pour RTL, près de 12 000 pour Europe 1 et 9 000 pour BFM TV), suivies de près par *Le Monde*, *Le Parisien*, *Le Point* ou encore *Le Figaro*. Les *pure players* et les journaux locaux (seuls *Le Dauphiné Libéré*, *La Provence*, *Nice Martin*, *Ouest France*, *Le Progrès*, *L'Union*, *Le Courrier Picard* et *Sud Ouest* – par ordre croissant de citations reçues – ont été cités plus de 1 000 fois en 2013) sont parmi les moins cités. Une première analyse des résultats laisse à penser que la prédominance des radios et des télévisions est majoritairement due à l'importance des interviews, et notamment aux reprises des matinales radios.

Au-delà de ces premières statistiques, que nous apprennent nos données sur les comportements de citation des médias ? Et en particulier, dans quelle mesure les médias créditent-ils davantage les concurrents qu'ils copient ? Pour répondre à cette question, nous avons utilisé l'une des particularités de notre algorithme de détection de copie qui consiste, pour chaque bloc de texte copié, à identifier l'article dans lequel il est apparu pour la première fois, et d'en attribuer ainsi la paternité au bon média. Ce qui nous permet de faire deux choses : d'une part, pour chaque article, d'identifier l'ensemble des médias qu'il copie (tout en quantifiant l'importance de la copie),

1 Il peut également parfois s'agir de médias étant les seuls en France à reprendre une information publiée à l'étranger. Ainsi, le site Pure People est le seul à noter le 30 janvier 2013 que Liberty Ross, « stylée et très sexy », semble bien vivre sa demande de divorce, et ce, « selon le New York Post, grâce à Jimmy Lovine [...] avec qui elle serait devenue très proche ». Sur une note plus sérieuse (et féministe !), l'article – non classé dans un événement – publié par le Huffington Post le 11 octobre selon lequel « les premiers artistes préhistoriques seraient des femmes » s'appuie sur une étude du *National Geographic* que le *pure player* cite explicitement : « Une récente étude montre que 75% de ces peintures auraient été faites par des femmes rapporte le *National Geographic*. »

2 Cependant, si un média est cité plusieurs fois à l'intérieur d'un article, nous ne comptabilisons qu'une seule citation. Il serait donc plus exact de dire qu'en moyenne un média est mentionné dans 4 500 articles publiés par des médias concurrents en 2013.

et d'autre part, de déterminer l'article (d'un autre média) dont provient la plus grande partie du contenu copié (si copie il y a). Ainsi, à l'intérieur de chaque événement et pour chaque article, nous savons précisément combien de fois un article a été repris comme « source primaire<sup>1</sup> », dans quelle mesure (le nombre de caractères repris) et par quel(s) média(s).

Nous pouvons ainsi mettre face à face la quantité du contenu d'un média qui a été copiée et le nombre de citations qu'il reçoit, et étudier la relation entre les deux à l'intérieur des événements. Dans seulement 8,7% des cas, les médias font référence au(x) média(s) concurrent(s) qu'ils copient, pourcentage qui s'effondre à 1,7% si l'on ne considère pas les mentions de l'AFP.

Les médias citent-ils uniquement le média concurrent qu'ils copient le plus? Si l'on se focalise sur le média le plus copié, on trouve que la probabilité qu'un média mentionne le concurrent qu'il copie le plus s'élève à 16,8%, mais retombe à 2,5% si l'on exclut à nouveau les citations à l'AFP. Cela pourrait provenir du fait que les médias préfèrent peut-être citer comme source de l'information le média *news breaker* plutôt que ceux qu'ils copient (dans le cas où ils diffèrent). Et de fait, le média *news breaker* est cité dans 18,6% des documents comme source de l'information. Toutefois, ce nombre tombe à 7,2% quand le *news breaker* n'est pas l'AFP.

Autrement dit, quelle que soit l'approche choisie, le crédit est loin – c'est peu dire – d'être une pratique très répandue. Au-delà des statistiques, un exemple qui aurait pu faire l'objet d'une mention dans le rapport de l'Observatoire de la Déontologie de l'Information : quand le 14 mars 2013 BFMTV reprend quasi en intégralité (le taux de copie est de 90%) une dépêche de l'AFP sur « des traces de porc trouvées dans des spaghettis bolognaises en Belgique », le site internet de la chaîne d'information en continu transforme un « contacté par l'AFP, Carrefour Belgique a indiqué... » en « contacté, Carrefour Belgique a indiqué... ». Autrement dit, fait sauter volontairement de son article toute référence à l'AFP.

<sup>1</sup> Nous avons fait le choix de tracer chaque bloc de texte jusqu'à sa première occurrence et de considérer comme copié le premier média à avoir publié ce bloc. Bien sûr, il est possible que dans certains cas un média A ait publié en premier une information, copié par le média B, puis que le média C ait ensuite copié ensuite le média B. Il nous est impossible de déterminer empiriquement la « source » utilisée par le journaliste du média C mais cela ne nous semble pas important. Ce qui compte ici, c'est de déterminer l'origine du texte copié, ce que nous pouvons en nous appuyant sur les heures précises de publication des différents articles.

Nous envisageons plusieurs pistes pour poursuivre l'étude de ces données et mieux caractériser les relations de citation. Cela passera notamment par l'identification des communiqués de presse, qui ne sont attribuables à aucun média, de l'« évaporation » des références au média d'origine comme source de l'information au fur et à mesure de la reprise des contenus et de la propagation de l'information, ainsi que par la caractérisation des citations liées aux diffusions d'interviews qui ne correspondent pas de fait à une copie, etc.

### **La production d'information originale explique-t-elle l'audience moyenne ?**

Nous avons vu précédemment que la production d'information originale n'influence que marginalement les parts d'audience des différents médias. Autrement dit, l'audience quotidienne d'un média ne varie que très peu en fonction de sa production d'information originale. Cela peut s'expliquer par la faiblesse de la pratique du crédit et des citations.

Mais nous avons également noté rapidement que l'audience moyenne, par exemple le nombre de visiteurs uniques, des différents médias varie fortement, de quelques dizaines de milliers de visiteurs à peine pour certains d'entre eux à plus d'un million pour d'autres. La production d'information originale n'expliquerait-elle pas ces différences de niveau ? Pour répondre à cette question, ce n'est pas aux variations quotidiennes de cette production qu'il faut s'intéresser ; il faut revenir à une analyse plus agrégée, par exemple au niveau de l'année.

En régressant l'audience moyenne des médias sur la quantité quotidienne moyenne d'information originale qu'ils produisent (en incluant les mêmes contrôles que ceux décrits précédemment), nous montrons que produire en moyenne un article original supplémentaire par jour augmente le nombre moyen de visiteurs uniques quotidiens de 14 000. Un journal comme *Le Monde* produit en moyenne l'équivalent de 32 articles entièrement originaux par jour, *Libération* l'équivalent de 15. D'après nos estimations, si *Libération* produisait une quantité équivalente à celle du *Monde*, alors son nombre de visiteurs uniques par jour augmenterait de 235 000. En d'autres termes, la seule quantité de contenu original produit (indépendamment par exemple du nombre de *breaking news*) explique 16% de la différence d'audience entre *Le Monde* et *Libération* en moyenne annuelle.

16%, le chiffre peut sembler relativement faible. Mais de nombreux facteurs entrent dans le choix de lecture des citoyens, des facteurs « subjectifs » – la tendance politique, les choix de traitement de la rédaction – et des facteurs plus « objectifs » – le nombre total d'articles, le nombre d'articles originaux, la diversité des sujets et des points de vue traités, des experts sollicités, la continuité rédactionnelle, etc. Si nous considérons que la réputation d'un média est une fonction de l'ensemble des critères objectifs, alors elle n'affecte pas seulement les fluctuations journalières de l'audience, elle joue également, à moyen et long terme, sur l'augmentation moyenne du nombre de visiteurs<sup>1</sup>. Si les médias n'ont donc pas « intérêt » à très court terme (du moins financièrement), à augmenter leur production de contenu original (avec les coûts supplémentaires que cela entraîne), produire davantage de contenu original semble néanmoins être une stratégie « rentable » à long terme, les gains de réputation permettant aux médias d'accroître leur audience moyenne.

### Les variations des ventes « hors ligne »

De plus, cet effet de réputation ne joue pas seulement « en ligne » mais également « hors ligne ». De manière anecdotique, la publication des « Panama Papers » aurait multiplié par trois les ventes en kiosque du *Monde* au cours de la semaine suivant la révélation du scoop. Certes, pour la plupart des médias (y compris *Le Monde*) les ventes en kiosque sont en chute libre et représentent un pourcentage de plus en plus faible de leurs abonnements. Mais cette diffusion continue néanmoins de représenter une part importante des revenus de la presse papier (et également une vitrine pour les sites internet).

Pour tester dans quelle mesure les ventes « hors ligne » varient avec la production d'information sur Internet, nous avons pour finir étudié les données de diffusion des journaux. Ces données ne sont disponibles qu'en moyenne mensuelle. Nous avons donc régressé la moyenne mensuelle de la diffusion des journaux sur leur production mensuelle moyenne d'information sur Internet.

La diffusion moyenne mensuelle d'un journal augmente avec la quantité d'information originale qu'il produit en ligne au cours du

<sup>1</sup> Pour valider totalement cette hypothèse, il faudrait suivre l'évolution des audiences en fonction des contenus sur une période très longue et comprenant des ruptures quant aux facteurs objectifs. En termes de données et de contexte, ce travail sort du cadre de notre étude.

mois (et qui est sans aucun doute, même si nous ne le mesurons pas directement ici, très fortement corrélée avec sa production d'information originale dans son édition papier). La magnitude de cet effet, si elle est faible, n'est pas négligeable. De plus, la diffusion mensuelle augmente également positivement avec le nombre de *breaking news*.

Certes, du fait de l'impossibilité d'utiliser des variations quotidiennes, ces effets ne sont pas bien « identifiés » et doivent être lus avec attention, bien plus comme des corrélations que comme des effets causaux. Ils reflètent néanmoins à nouveau l'importance jouée à moyen terme par les effets de réputation. On peut regretter que de plus en plus nombreux soient les actionnaires de médias – on pense en particulier à Patrick Drahi – aient une vision entièrement court-termiste et comptable de ce secteur d'activité. Pour eux, l'information a un prix – celui de son coût de production – et celui-ci doit rester inférieur aux bénéfices qu'ils espèrent en retirer. Ce n'est pas notre vision de l'information, il s'agit pour nous d'un bien public. Dans le contexte actuel de montée des populismes, il est urgent de préserver la réputation des médias et de reconstruire un lien de confiance entre les médias et les citoyens. À tout prix.

### **Un nouvel enjeu pour les médias : garder les utilisateurs dans leur environnement**

L'un des enjeux principaux aujourd'hui pour les médias – ceux qui le souhaitent, une autre stratégie pour des petits médias « parasites » étant de se comporter en passager clandestin, ce qui pourrait aboutir à une polarisation croissante de la taille des rédactions – est donc de se construire une réputation telle que les utilisateurs souhaitent consommer ces médias (et les consommer « en direct », sans passer par les intermédiaires de l'information en ligne que l'on appelle « infomédiaires<sup>1</sup> », agrégateurs de contenu ou réseaux sociaux) plutôt que de choisir indifféremment entre toutes les sources d'information disponibles. Cet enjeu est de taille car, sur Internet, les citoyens semblent de moins en moins « engagés » par rapport aux médias qu'ils consomment.

<sup>1</sup> Sur les infomédiaires, le lecteur intéressé pourra se reporter à l'excellent livre de Frank Rebillard et Nikos Smyrniaios : Rebillard, F. et Smyrniaios, N. (2010). *Les infomédiaires, au cœur de la filière de l'information en ligne*, Paris, La Découverte.

En 2013, la durée moyenne des visites sur les sites internet des médias était ainsi inférieure à quatre minutes, les utilisateurs ne consultant en moyenne que trois pages (et passant donc moins d'une minute par page<sup>1</sup>). De plus, le « taux de rebond » ou *bounce rate* – taux qui mesure le pourcentage d'internautes qui, une fois rentrés sur un site internet, le quittent sans consulter d'autres pages – était de 56%. Autrement dit, plus de la moitié des internautes quittaient le site internet des médias après avoir consulté seulement une page.

Un nouveau défi pour les médias est donc d'arriver à conserver les lecteurs à l'intérieur de leur environnement numérique. Il est d'ailleurs intéressant de ce point de vue qu'un média comme *Mediapart*, qui a fait depuis son lancement le choix du tout payant, faisant le pari des contenus à haute valeur ajoutée plutôt que des flux d'information, ait décidé d'incorporer sur son site un fil Reuters d'actualités automatisé.

*Mediapart* a également fait le choix de multiplier les soirées spéciales en direct, qui se présentent – et sont organisées – comme des émissions de télévision visibles sur Internet. *Le pure player* est d'ailleurs loin d'être le seul à avoir fait en ligne le pari du format de la télévision, puisque des journaux comme *Le Monde* ou *Le Figaro* donnent de plus en plus à voir à leurs lecteurs de courtes émissions en plateau tournées directement à l'intérieur des rédactions.

---

1 Ces données proviennent de SimilarWeb.



## CHAPITRE 7

### QUELLES SOLUTIONS ?

Nous nous retrouvons donc face à une contradiction majeure : d'une part l'information, du fait des coûts fixes que représentent les journalistes, coûte cher à produire ; d'autre part, du fait de sa vitesse élevée de propagation et de la gratuité en ligne, elle rapporte de moins en moins aux médias qui sont ainsi moins incités à en produire.

Tout n'est pas noir pour autant. Tout d'abord, comme nous venons de le voir, les médias ont encore une incitation majeure à produire de l'information de qualité : leur réputation.

De plus, s'il y a des raisons d'être pessimiste – notamment lorsque l'on s'aperçoit de la diminution du nombre de journalistes un peu partout dans les pays développés, à commencer par la France – il y a aussi un grand nombre de raisons d'être optimiste. Le paysage médiatique français est en plein bouillonnement. Chaque jour, de nouveaux acteurs apparaissent et de nouveaux modèles voient le jour, sur le papier comme en ligne.

*Un paysage médiatique en bouillonnement...*

Sur Internet, une seconde génération de *pure players* a fait son apparition, délaissant l'actualité « chaude » pour le journalisme « lent » et la production d'une information à haute valeur ajoutée. On peut citer – parmi beaucoup d'autres – *Les Jours* créé par d'anciens journalistes de *Libération*, *Le Quatre Heures* qui offre à lire une seule enquête par mois, *Ijsberg*, *L'imprévu* dont le principe éditorial est de revenir sur les événements oubliés, ou encore *8<sup>e</sup> étage* dont le slogan est « différencier l'information de l'actualité<sup>1</sup> ».

---

1 Ce bouillonnement médiatique est bien sûr loin d'être propre à la France. Les nouveaux



On observe également le développement rapide d'un certain nombre de sites d'information spécialisés, comme Contexte, spécialisé dans les politiques publiques. Ces sites – qui d'une certaine manière s'inscrivent dans la lignée des lettres d'information d'Indigo Publications –, de par leurs tarifs extrêmement élevés, ne s'adressent pas au grand public, mais ils ne peuvent néanmoins être ignorés dans l'analyse de la production d'information sur Internet. Au contraire, Brief.me, newsletter quotidienne d'information généraliste, s'adresse, elle, au grand public à qui elle propose de faire gagner du temps en hiérarchisant l'information pour ses lecteurs.

Sur le papier également, même si cela ne nous concerne pas directement ici, de nouveaux médias ont émergé avec succès ces dernières années, comme *Le I* – à l'équilibre seulement deux ans après sa création –, *Society*, ou encore récemment à Paris *Soixante-Quinze*, un magazine de société consacré à la ville.

*... qui témoigne tout autant d'une grande vitalité que d'une grande fragilité*

Tout n'est pas rose pour autant. Les nouveaux *pure players* français en particulier ont des modèles économiques qui restent extrêmement fragiles. D'un côté, ceux qui ont fait le choix de la publicité s'aperçoivent qu'il est extrêmement difficile de monétiser son audience en ligne, et ce encore plus quand cette audience est faible<sup>1</sup>. Ceux qui ont fait le choix du payant s'en sortent relativement mieux aujourd'hui, mais fidéliser suffisamment d'abonnés pour devenir rentable prend du temps et les rédactions de ces nouveaux médias restent encore aujourd'hui en moyenne de taille extrêmement réduite. Les conditions de travail peuvent y être également difficiles, comme en témoignent les débats récurrents sur la rémunération des pigistes<sup>2</sup>. Beaucoup de ces nouveaux médias – sans doute par nécessité plus que par choix – rémunèrent d'ailleurs leurs contributeurs

---

médias et les nouveaux formats sont légion, notamment aux États-Unis. Certains s'importent d'ailleurs en France, comme en témoigne le lancement de *Mashable France* en partenariat avec France 24, avec 9 journalistes et un budget d'environ un million d'euros par an. Sans oublier bien sûr le lancement de *Vice France* (le *Vice News* français) en association avec *Les Inrockuptibles* et Radio Nova.

1 De plus, le recours croissant à la publicité native pose des problèmes de déontologie.

2 Ou plutôt sur leur « non rémunération », dénoncée par exemple par Acrimed dans le cas de *Ijsberg*.

comme autoentrepreneurs et non en piges, ce qui empêche parfois ces contributeurs d'obtenir leur carte de presse<sup>1</sup>.

De plus, si certains de ces nouveaux médias payants ont réussi à devenir rentables – on pense notamment à *Mediapart* – la question clef à laquelle il faut répondre aujourd'hui est la suivante : pour combien de ces nouveaux médias les Français sont-ils prêts à payer ? Le succès de *Mediapart* (entre autres) a poussé certains de ces nouveaux acteurs à fixer le prix de leur abonnement à 9€ par mois. Mais ce n'est pas seulement la somme que les citoyens ont été prêts à déboursier pour un abonnement à *Mediapart*<sup>2</sup>, c'est la somme *totale* qu'ils sont en moyenne prêts à dépenser chaque mois sur le poste de dépenses « information ». En d'autres termes, il n'est pas sûr qu'il y ait suffisamment « de place » pour tous ces nouveaux acteurs.

Or, la plupart d'entre eux, du fait de leur taille, ne sont pas généralistes, ou, quand ils le sont, ne sont pas exhaustifs. Autrement dit, s'informer de manière exhaustive pour un citoyen suppose de s'abonner à plusieurs d'entre eux, et se repose la question des montants que les Français sont prêts à déboursier. De plus, l'histoire récente américaine montre qu'il existe un véritable risque qu'un certain nombre de ces *pure players* renonce dans un futur proche au moins en partie à produire de l'information de qualité (coûteuse) et décide au contraire de faire le choix du divertissement, tout à la fois moins cher à produire et plus rentable en termes de publicité (**ENCADRÉ**).

### **BuzzFeed, Mashable... : déjà la fin de l'information ?**

On a vu apparaître ces dernières années aux États-Unis un grand nombre de sites qui, après s'être lancés principalement sur le créneau du divertissement (« *entertainment* »), ont commencé à investir dans la production d'information et à (dé)embaucher des journalistes de médias traditionnels.

1 Selon la loi, tout journaliste doit être rémunéré en salaire (au mois ou à la pige). Les paiements en droits d'auteur ou en honoraires ainsi que le statut d'autoentrepreneur sont ainsi considérés contradictoires avec le statut de journaliste professionnel.

2 Le prix de l'abonnement à *Mediapart* est depuis avril 2016 passé à 11€ par mois.

L'équipe d'investigation de l'édition anglaise de BuzzFeed (qui emploie au total 170 journalistes) compte ainsi aujourd'hui 20 journalistes, dont d'anciens de *Politico* ou du *Financial Times*. *Gawker* emploie 132 journalistes quand *Mashable*, jusqu'au début de l'année 2016, en employait 70. Mais *Mashable*, après seulement quelques années d'existence, vient de réduire à 40 le nombre de ses journalistes afin de se recentrer sur la vidéo. Pourquoi? Parce que la vidéo est beaucoup plus rentable (du moins à court terme) que la production d'information.

Le même virage s'annonce du côté de *BuzzFeed* qui a raté ses objectifs pour 2015 (170 millions de dollars de chiffre d'affaires contre des objectifs de 210 à 250 millions). *BuzzFeed* a ainsi annoncé à l'été 2016 qu'il allait séparer ses activités d'information d'une part (*news division*) et ses activités de divertissement (*entertainment division*), en particulier la production de vidéos, de l'autre. Ce qui ne laisse rien augurer de bon quant aux ressources qui seront consacrées dans le futur à la production d'information.

On pourrait argumenter qu'aujourd'hui, l'information généraliste est disponible de manière exhaustive gratuitement en ligne, et que les citoyens qui s'abonnent à ces nouveaux médias ne le font pas pour avoir accès à une information qu'ils trouvent gratuitement ailleurs, mais justement pour consommer une information différente, à haute valeur ajoutée, pour laquelle ils sont prêts à payer. Soit. Mais l'on en revient alors à la question de départ. Qui produit – et qui paie – pour la production d'une information généraliste, exhaustive et gratuite en ligne?

Pourquoi les médias semblent-ils moins incités aujourd'hui à produire de l'information originale qu'ils ne l'étaient par le passé? Parce qu'ils arrivent de moins en moins à monétiser cette information. Autrement dit, ils ne perçoivent pas de revenus suffisants pour compenser le coût de la production. Une partie de la solution réside dans le fait d'augmenter les revenus liés à la production d'information originale, ce que permettent les murs payants ou la syndication de contenu. Une autre partie de la solution réside dans le fait de dimi-

nuer les coûts de production de l'information, par exemple en mutualisant ces coûts entre les différents médias. Ce chapitre discute ces différentes solutions, sans oublier bien sûr la nécessité d'un soutien public à la production de ce bien public qu'est l'information.

### **Les murs payants : une solution déjà mise à mal ?**

Le *New York Times* a vu ses revenus publicitaires s'effondrer au deuxième trimestre 2016, non seulement sur papier (-14,1%) mais également en ligne (-6,8%). Doit-on pour autant s'inquiéter de la mauvaise santé économique du quotidien ? La réponse peut sembler surprenante mais elle est négative. Le *New York Times* se porte plutôt bien, comme en témoigne la relative stabilité de son chiffre d'affaires. Cette stabilité s'explique par la hausse tout à la fois des ventes en kiosque et des ventes par abonnement, papier et en ligne. Le *New York Times* compte aujourd'hui près de 1,5 million d'abonnés en ligne uniquement. Les recettes de ce succès ? L'introduction d'un mur payant (*paywall*).

Face à l'effondrement des recettes publicitaires des médias – lié en grande partie à la capture du marché publicitaire numérique par un nombre d'acteurs extrêmement réduit, et en particulier Facebook qui accapare à lui seul 30% des revenus publicitaires sur Internet aux États-Unis – les murs payants apparaissent aujourd'hui comme le seul modèle économique viable sur le long terme pour monétiser son audience. Le *paywall* est d'ailleurs devenu largement la norme aux États-Unis où, d'après l'American Press Institute, 78% des journaux dont la diffusion dépasse les 50 000 exemplaires, en ont mis un en place.

En France – alors que cela était encore très minoritairement le cas en 2013 – les principaux médias sont passés petit à petit derrière un mur payant, qu'il s'agisse des *Échos*, du *Monde* ou encore du *Figaro* mais également de la presse quotidienne régionale (*La Dépêche du Midi*, *La Voix du Nord*, ou encore *Sud Ouest*).

#### *Le Sun, crépuscule des murs payants ?*

Face au développement un peu partout des murs payants, certains opposent l'exemple du *Sun*, qui témoignerait de l'échec des murs payants. Mais que nous apprend véritablement le cas du *Sun* ?

Le *Sun*, qui appartient à Rupert Murdoch, a renoncé à son *paywall* en novembre 2015, *paywall* qu'il avait introduit seulement deux ans auparavant et qui avait conduit à un effondrement de son audience en ligne. Mais cela ne témoigne pas tant de l'échec du mur payant, que de l'échec du payant pour un média qui n'offre pas de contenu à forte valeur ajoutée (mais des potins de stars et de sportifs).

Pour que les lecteurs soient prêts à payer, il faut leur donner une raison de payer, leur offrir un contenu à haute valeur ajoutée qu'ils ne peuvent pas trouver ailleurs gratuitement. Personne ne s'abonnera à un média si l'information qui s'y trouve ressemble à un fil Twitter. Le *Times* de Londres (qui est derrière un *paywall* étanche) l'a d'ailleurs bien compris et c'est pourquoi il a décidé au début de l'année de renoncer au flux d'informations et de ne plus actualiser son site internet que trois fois par jour. On l'a vu, la production d'information se fait sous contrainte de ressources. Renoncer au fil d'informations, c'est consacrer de nouvelles ressources à des analyses fouillées.

Le *New York Times*, lui, n'a pas annoncé renoncer à l'actualité chaude, mais il a néanmoins pris la décision de consacrer moins de ressources dans le futur aux « *commodity news* » et davantage à ses contenus propres à haute valeur ajoutée : « *only-in-The New York Times* », choix cohérent avec le modèle du payant<sup>1</sup>.

### *Instant Articles : la fin des paywalls ?*

La question du « gratuit ou payant » est-elle pour autant tranchée définitivement ? On aurait pu le penser jusqu'au lancement en 2015 d'Instant Articles par Facebook qui est venu fragiliser l'équilibre déjà précaire d'un certain nombre de médias. Instant Articles, c'est la publication par les éditeurs de leurs articles directement sur Facebook qui devient ainsi hébergeur d'articles de presse<sup>2</sup>. En d'autres termes, les médias qui font le choix de publier du contenu sur Instant Articles voient leurs lecteurs le consommer sur Facebook et non plus sur leur propre application. Ce qui suppose que ce contenu soit disponible gratuitement sur Facebook et non plus payant sur l'application de l'éditeur.

1 Arthur Sulzberger, alors éditeur du *New York Times*, avait d'ailleurs déclaré en 2011 (le 6 avril à la Columbia Journalism School) que ce que le journal devait viser, c'était de l'information de qualité pour une audience de qualité (« *quality news for quality audiences* »).

2 Jusqu'à présent, les éditeurs ne publiaient sur Facebook que des liens vers leurs articles, pas les articles eux-mêmes.

Au-delà de la question sensible de la gestion de la publicité et du partage de ses revenus, tout l'enjeu est de savoir si cela ne fragilisera pas à terme les murs payants. Certains éditeurs (comme le *New York Times*) ont fait le choix pour l'instant de ne publier en Instant Articles qu'un nombre réduit d'articles par jour, ce qui est cohérent avec la logique de leur *soft paywall*. La plupart des médias – à l'exception de ceux qui sont derrière un *hard paywall* – offrent en effet déjà à lire un certain nombre d'articles gratuitement, comme une sorte de produit d'appel<sup>1</sup>. Pourquoi dès lors ne pas le faire sur Facebook qui, on le sait, génère un trafic extrêmement important? Le risque, bien réel, c'est qu'Instant Articles soit appelé à se développer et à prendre une importance telle que la majorité des utilisateurs consommeront uniquement leur information à partir de Facebook. Les éditeurs ne seront-ils pas alors amenés à y publier, « contre leur volonté », une quantité toujours plus importante de contenu *gratuit*? De plus, en entraînant une baisse de trafic sur les sites internet des éditeurs, Instant Articles rend plus compliqué pour ces derniers de convaincre les utilisateurs de s'abonner<sup>2</sup>.

Facebook n'est pas le premier à ralentir les médias dans leur volonté de faire payer les lecteurs. Avant lui, Google a empêché pendant des années aux médias derrière un mur payant d'avoir leur contenu bien indexé, notamment ceux dont le *paywall* était *hard* (aucun article n'étant en accès libre). Une autre menace qui pèse sur les murs payants, c'est le non-respect de la propriété intellectuelle.

Même un journal comme le *Wall Street Journal*, journal de référence s'il en est un quant au succès du mur payant – le titre compte aujourd'hui près de 950 000 abonnés numériques exclusifs – est en train de repenser son modèle de mur payant. Il essaie d'introduire un *paywall* plus souple, en proposant des offres permettant aux non-abonnés de tester le journal comme invités pendant 24 heures, offre qui apparaît lorsqu'un lecteur accède à un article partagé par un abonné ou un journaliste<sup>3</sup>.

1 En France, *Le Parisien*, *20 Minutes* et *Les Échos* ont été les premiers à publier des contenus sur ce nouveau format, tout comme *Libération*.

2 Les mêmes problématiques se posent avec l'application « News » d'Apple. Cette application – dans l'esprit d'Instant Articles – permet d'offrir une sélection personnalisée d'articles de médias partenaires et surtout d'héberger une partie de ces contenus.

3 Voir par exemple l'article de Shan Wang publié le 15 août 2016 sur le site du NiemanLab : « The Wall Street Journal is changing up its paywall, offering guest passes and expanded link-sharing on social. » <<http://www.niemanlab.org/2016/08/the-wall-street-journal-is-changing-up-its-paywall-offering-guest-passes-and-expanded-link-sharing-on-social/>> Le *Wall Street*

Techniquement, il reste compliqué pour les médias de mettre en place de façon entièrement efficace des murs payants au compteur, les fameux « *metered paywalls* » – aussi appelés parfois Purposefully Porous Paywall (PPP) – car les citoyens consomment de plus en plus l'information sur différents supports, c'est-à-dire avec différentes adresses IP. Ces murs payants au compteur sont de plus contournables par les réseaux sociaux – il suffit par exemple de commencer par tweeter un article que l'on veut lire puis de cliquer sur le lien Twitter renvoyant vers l'article – ainsi qu'en supprimant les cookies de navigation ou encore en utilisant la navigation privée sur les navigateurs les plus récents. Nous défendons dans ce livre l'idée selon laquelle les citoyens devraient davantage payer pour l'information qu'ils consomment, et nous ne cherchons donc nullement à inciter à l'utilisation de toutes ces méthodes de contournement. Mais devrait-on pour autant ignorer leur existence ? Pouvoir supprimer les cookies de navigation est par exemple un élément clef de la protection de la vie privée sur Internet et ne devrait pas rentrer en contradiction avec les murs payants.

*Avant le mur payant, penser au copyright ?*

Imaginons le cas où deux médias sont en concurrence, l'un derrière un mur payant et l'autre en accès libre. Le modèle économique du média payant semble le bon, pour les raisons que nous avons évoquées plus haut. Le mur payant lui permet de générer des revenus d'abonnement, quand son concurrent monétise difficilement son audience grâce à la publicité. Imaginons maintenant que le média gratuit publie sur son site internet du contenu directement repris (sous la forme de copié-collé) du site concurrent du média payant. Le lecteur peut accéder au même contenu soit gratuitement sur le site du média gratuit, soit en payant sur le site du média payant. Tout laisse à penser qu'il ne choisira pas dans ce cas de s'abonner au média payant, et ce dernier n'aura donc aucune incitation à continuer à payer des journalistes pour produire du contenu original. En d'autres termes, en l'absence de respect du copyright, l'introduction de murs payants n'est-elle pas une fausse bonne solution ?

---

*Journal* qui n'échappe d'ailleurs pas à la crise, le titre ayant annoncé une refonte de son édition papier en novembre 2016, afin en particulier de pouvoir réduire les dépenses, autrement dit le nombre de journalistes.

L'exemple que nous venons de prendre peut sembler irréaliste, pourtant il ne l'est pas, comme en témoigne le cas récent de deux journaux sud-africains, *Moneyweb* et *Fin24*, qui ont porté leur conflit devant les tribunaux en 2013.

*Moneyweb* a en effet attaqué en justice *Fin24* pour violation du copyright après avoir constaté que *Fin24* avait copié de manière extensive plusieurs de ses articles et en avait publié des versions retravaillées sur son site internet. Selon *Moneyweb*, ces articles originaux avaient demandé beaucoup d'efforts et de ressources pour être produits. *Fin24*, en faisant le choix d'une stratégie que l'on pourrait qualifier de « low cost » – en recopiant non seulement les faits mais également l'expression journalistique des articles de *Moneyweb* – aurait affaibli les incitations de ce dernier à produire une information originale et à employer des journalistes. *Moneyweb* a donc attaqué *Fin24* non seulement au titre de la violation du copyright (Copyright Act 98 de 1978) mais également de la concurrence déloyale.

De manière intéressante, dans ses arguments développés tout au long du procès, *Fin24* n'a pas nié avoir copié des articles de *Moneyweb*, mais a fait référence à trois aspects importants des lois sud-africaines sur la propriété intellectuelle : d'une part, l'originalité ; d'autre part, la substantialité ; et enfin le fait que d'après la section 12(8)(a) du Copyright Act 98, il n'y aurait pas de protection du copyright concernant les nouvelles du jour quand l'information utilisée dans les articles provient d'une tierce partie, cette tierce partie pouvant comprendre par exemple des communiqués de presse ou des interviews<sup>1</sup>. Pour qu'un article soit protégé par le copyright, il faut qu'il ait fait l'objet d'un travail « original » et créatif ; or la seule présence d'un journaliste prenant des notes à une conférence de presse ne saurait constituer un tel travail selon une certaine interprétation de ce principe d'originalité. Autre argument utilisé par *Fin24* et sur lequel nous reviendrons à la fin de ce chapitre : la libre circulation de l'information est un droit constitutionnel. Autrement dit, en reprenant les articles de *Moneyweb* et en diffusant une version remaniée sur son site internet, *Fin24* répondait à une approche de l'information comme « bien public » dont on peut souhaiter que le plus grand nombre y ait accès.

La Haute Cour du Gauteng a rendu sa décision en mai 2016. Alors que *Fin24* était mis en accusation concernant sept articles de *Mo-*

<sup>1</sup> « No copyright shall subsist [...] in news of the day that are mere items of press information. »

*neyweb*, elle a décidé que le site internet n'avait enfreint la loi sur le copyright qu'une seule fois. Il est intéressant de noter à quel titre d'après la Cour le respect du copyright n'a pas été enfreint dans les autres cas. En particulier, concernant quatre des sept articles mis en cause par *Moneyweb*, la Cour a considéré que *Moneyweb* n'avait pas prouvé que ces articles étaient le fruit d'un travail original, et qu'ils n'étaient donc pas protégés par la loi sur le copyright. Concernant le premier article par exemple, écrit par une journaliste de *Moneyweb* après qu'elle ait participé à une conférence de presse au Parlement, le juge a décidé qu'il n'y avait pas assez de preuves permettant d'établir que son travail était original, le contenu de l'article reposant essentiellement sur des informations rendues publiques (à tous) lors de la conférence. Le juge a pris exactement la même décision concernant un article écrit par une journaliste ayant participé à une conférence téléphonique – avec des journalistes d'autres médias – avec le dirigeant d'une entreprise, ou encore un article écrit par une journaliste qui avait participé – avec deux autres journalistes travaillant pour des groupes de médias concurrents – à la visite de presse d'un restaurant. Dans ces trois cas, les informations révélées (lors de la conférence de presse, de l'entretien téléphonique ou de la visite) l'ont été alors que les émetteurs de l'information savaient qu'elle était ainsi mise dans le domaine public.

Au-delà de ce cas d'école qui risque d'impacter des décisions futures sur la question de la propriété intellectuelle dans les médias, quels sont les avantages et les limites du copyright? Celui-ci représente-t-il une solution de long terme au problème du manque d'incitations financières dont souffre aujourd'hui la production d'information sur Internet?

### **Avantages et limites du copyright**

Traditionnellement, une violation du copyright est constatée lorsqu'une personne ou une organisation recopie et diffuse des passages conséquents d'une œuvre produite par autrui en son nom.

Il existe, en matière de médias, différents fondements juridiques. En France, les droits de l'auteur contre la contrefaçon sont protégés par le Code de la propriété intellectuelle. Le journaliste est protégé en tant qu'auteur de son article de la même manière qu'un écrivain

est protégé pour le contenu de son livre<sup>1</sup>. Le fait pour un média de reprendre in extenso le contenu original d'un autre média constitue donc une contrefaçon.

*L'absence de protection des faits*

Il est cependant important de noter que le droit français de la propriété intellectuelle ne protège ni les idées, ni les faits historiques. Ce droit est en effet fondé sur le principe que « la propriété littéraire et artistique ne protège pas les idées ou concepts, mais seulement la forme originale sous laquelle ils se sont exprimés<sup>2</sup> ». De même, il a été admis que « les faits historiques ou purement biographiques, ou les informations ne peuvent en eux-mêmes subir la moindre appropriation<sup>3</sup> ». Le droit confère donc une protection à l'œuvre, à sa composition et à son expression, mais non à l'idée qui est à son origine. Il ne protège que la forme originale sous laquelle les idées sont présentées.

De même, en droit américain, une idée ne peut pas être copyrightée (pas plus que des données). La loi américaine souligne en effet que, pour être protégé, un travail doit faire l'objet d'une « expression originale ». Autrement dit, un fait ne peut pas faire l'objet d'une protection : il ne peut pas être copyrighté. Seule la façon dont une personne exprime des idées ou des faits peut l'être<sup>4</sup>. À une exception notable près : la doctrine des « *hot news* » ou « *misappropriation doctrine* » qui spécifie que l'on peut considérer qu'un fait est « approprié » de manière abusive s'il est reproduit dans un délai trop court de sorte que le producteur de l'information n'a pas le temps d'en tirer un bénéfice suffisant pour compenser ses coûts (**ENCADRÉ**<sup>5</sup>).

1 On distingue le droit d'exploitation de l'article qui appartient au journal (à l'éditeur) (article L. 122-1 CPI) du droit moral sur l'œuvre qui reste nécessairement la propriété de l'auteur, c'est-à-dire du journaliste (de l'éditeur) (article L. 121-1 CPI). Nous y reviendrons dans la section sur le droit voisin.

2 Voir en particulier la décision de la Chambre Civile 1 de la Cour de Cassation du 29 novembre 2005 (n°04-12.721).

3 Selon la décision du Tribunal de Grande Instance de Paris, 3e chambre 2e section, du 26 novembre 2010.

4 La jurisprudence américaine relève ainsi « le caractère moins protégé des œuvres d'information par rapport aux œuvres de l'imagination » (Ginsburg, J. et Sirinelli, P. (2010). « Google Book Search. Les enjeux internationaux pour le droit d'auteur », *La Semaine juridique*, 26 avril 2010).

5 Sur la doctrine des « *hots news* », le lecteur intéressé pourra notamment se reporter à Balganesch, S. (2011). « "Hot News" : The Enduring Myth of Property in News », *Columbia Law Review*, 111(3): 419-497.

### **La doctrine des « *hots news* » : de la possibilité de protéger (pour un temps) un fait**

La doctrine des « *hot news* » a été mise en place suite à une décision de la Cour suprême américaine en 1918 suite à un différend entre deux agences de presse, Associated Press (AP) d'une part et l'International News Services (INS) de l'autre.

Située sur la côte ouest des États-Unis, l'INS se contentait de reprendre les faits diffusés par les journaux de la côte est suite aux dépêches d'AP (en allant parfois jusqu'à corrompre les employés des journaux afin d'avoir accès au journal avant sa publication!) et – tirant profit du décalage horaire – vendait ses dépêches à des tarifs plus avantageux que ceux d'AP aux journaux de la côte ouest. (L'INS pouvait ainsi tirer un bénéfice élevé d'un investissement relativement faible; notons toutefois que son propriétaire d'alors, William Randolph Hearst, du fait de ses fortes critiques de la Grande Bretagne et de l'entrée en guerre des États-Unis ainsi que de ses amitiés allemandes, n'avait pas vraiment le choix puisque la Grande Bretagne avait interdit aux reporters d'INS d'envoyer des télégrammes sur la guerre.) La Cour suprême a considéré qu'il s'agissait là de concurrence déloyale qui ne permettait pas à AP de retirer un bénéfice des dépenses générées pour mettre à jour ces faits, en particulier dans le cadre de la couverture de la Première Guerre mondiale en Europe.

En France, quelques années plus tôt, des condamnations similaires étaient intervenues sur ce même fondement de la concurrence déloyale. Ainsi, une décision de la Chambre des requêtes du 23 mai 1900 concernant l'Agence Havas note que « les dépêches et nouvelles de l'Agence Havas [...] constituent [...] une propriété particulière acquise à grands frais et conférant à celle-ci et à ses abonnés un droit exclusif à la propriété de la publication jusqu'au moment où, soit par son fait, soit par celui de ses abonnés, elles ont été mises en circulation et sont tombées dans le domaine public ».

Afin de s'assurer qu'AP continue à fournir les investissements nécessaires à la production d'information, la Cour suprême a exigé de l'INS qu'elle attende un délai suffisant avant de publier les faits afin de permettre à AP de pouvoir retirer un bénéfice de son investissement dans l'intervalle. La portée de cette doctrine a toutefois été réduite par le Copyright Act de 1976. Si elle existe toujours (elle a par exemple été utilisée en 2009 par Associated Press : AP vs. All Headline News Corp.), son utilisation doit être revue à la lumière des changements technologiques de ces dernières années. Comment protéger un fait – une « *hot news* » – suffisamment longtemps à l'ère de la diffusion instantanée de l'information et des agrégateurs automatiques de contenus ?

Cette distinction entre la protection de l'expression et l'absence de protection des faits et des idées est essentielle quand l'on s'intéresse plus précisément à la protection accordée aux auteurs d'articles de presse. En effet, c'est suivant cette logique que le droit français a distingué – de longue date<sup>1</sup> – « les articles de fond, politiques et littéraires publiés par les journaux » qui sont considérés comme des œuvres de l'esprit et le produit du travail des journalistes, des « simples nouvelles<sup>2</sup> ».

L'une des difficultés aujourd'hui et l'un des enjeux pour les réflexions futures sur la propriété intellectuelle est d'ailleurs la question de la protection des articles de presse « écrits » par les journalistes « robots » que nous avons évoqués au chapitre 3.

1 Cette jurisprudence résulte d'un arrêt de la Chambre des requêtes de la Cour de Cassation du 8 août 1861 dans l'affaire Havas (ancêtre de l'AFP).

2 La Convention de Berne du 9 septembre 1886 pour la protection des œuvres littéraires et artistiques (traité diplomatique qui établit les fondements de la protection internationale des œuvres et qui comptait 168 parties contractantes en 2015) précise de même à l'article 2.8 : « La protection de la présente convention ne s'applique pas aux nouvelles du jour ou aux faits divers qui ont le caractère de simples informations de presse. »

*L'exception de courte citation*

Au-delà des simples nouvelles, un média peut reproduire le contenu d'un autre média dans le cadre de l'exception de courte citation. Cette exception est toutefois – c'est important – soumise à l'obligation de citer la source du passage de l'article qui est reproduit<sup>1</sup>.

Le système des exceptions est dit « fermé » en France. Comme le soulignent Ginsburg et Sirinelli (2010), « la liste des exceptions est donnée [...] de manière très descriptive (cas précis) et limitative. Les juges doivent s'en tenir à cette lecture restrictive et ne pas déborder le cadre des exceptions admises<sup>2</sup> ».

Aux États-Unis, au contraire, le système des exceptions peut être vu comme beaucoup plus « ouvert » puisque la loi américaine sur le copyright (le Copyright Act de 1976<sup>3</sup>) admet une dérogation générale au droit d'auteur susceptible d'être mise en œuvre dans n'importe quelle situation (il n'existe pas de cas prédéterminés). L'article 107 du Copyright Act introduit en effet l'exception de « fair use » au monopole de l'auteur et « oblige les juges confrontés au moyen de défense invoquant le fair use à prendre en considération les facteurs suivants : 1. les buts et les caractéristiques de l'utilisation, y compris si l'utilisation est de nature non commerciale ou poursuit des buts d'éducation non lucrative; 2. la nature de l'œuvre protégée; 3. la quantité et le caractère substantiel de la partie utilisée par rapport à l'ensemble de l'œuvre protégée; 4. l'incidence de l'utilisation sur le marché potentiel ou la valeur de l'œuvre protégée<sup>4</sup> ». Ainsi, cette approche est beaucoup plus souple que la seule exception envisagée par le droit français, celle de « courte citation ».

En matière de nouvelles techniques, la jurisprudence américaine indique également que « si les juges craignent que le respect du droit

1 « Lorsque l'œuvre a été divulguée, l'auteur ne peut interdire : [...] 3° sous réserve que soient indiqués clairement le nom de l'auteur et la source : a) Les analyses et courtes citations justifiées par le caractère critique, polémique, pédagogique, scientifique ou d'information de l'œuvre à laquelle elles sont incorporées » (article L. 122-5 CPI).

2 La question du respect de la propriété intellectuelle dépasse bien sûr de très loin la seule question des médias et le lecteur intéressé par ces questions pourra se référer aux développements extrêmement intéressants de Ginsburg et Sirinelli (*op. cit.*), en particulier en ce qui concerne la numérisation non autorisée de livres et leur stockage par Google.

3 Voir par exemple Fox, A. (2009). « Copyright, Competition and Publishers' Pursuit of Online Compensation », Technical report. USC Annenberg Center on Communication Leadership & Policy.

4 Ginsburg et Sirinelli, *op. cit.* Notons de plus que les juridictions américaines n'imposent pas la réunion stricte de ces quatre conditions.

d'auteur entraîne le blocage des bienfaits de la technologie, ils vont jouer le *fair use* pour débloquer la situation ». Cet argument a parfois joué en faveur de Google dans des conflits concernant le respect de la propriété intellectuelle, et notamment dans le cas de la numérisation des livres, Google soutenant que l'obtention d'autorisations de chaque ayant droit serait trop coûteuse et freinerait la mise à disposition et l'indexation de livres du monde entier.

### *Le droit d'auteur et les agrégateurs*

À l'exception de l'exemple sud-africain évoqué précédemment, les principales batailles juridiques menées par les médias au nom du droit d'auteur l'ont été contre des agrégateurs, en particulier Google News. Si Google News est hors du champ direct de cette étude (même si le fonctionnement de son algorithme n'est pas étranger aux pratiques de publication en ligne de nombreux médias), il est toutefois intéressant de se pencher un instant sur son mode de fonctionnement.

Google News présente de façon automatisée des articles d'information en provenance de sites internet. Cela a donné lieu à de nombreux débats – et batailles juridiques – au nom de la violation du copyright (**ENCADRÉ**). Pour sa défense, Google utilise tout d'abord l'exception de courte citation que nous venons d'évoquer. En droit français comme en droit américain, les titres ou de courtes citations ne sont en effet pas protégés par les lois sur le copyright. Or Google ne présente sur son site d'actualités que le titre, le chapeau et les photos illustrant les articles ; il ne publie jamais les articles dans leur intégralité.

N'y a-t-il aucun autre fondement juridique opposable à Google News ? Dans l'esprit de la doctrine du « *hot news* » américaine et dans le cadre du droit français, on pourrait lui opposer le parasitisme « caractérisé par la circonstance selon laquelle une personne, à titre lucratif et de façon injustifiée, s'inspire ou copie une valeur économique d'autrui, individualisée et procurant un avantage concurrentiel, fruit d'un savoir-faire, d'un travail intellectuel et d'investissement<sup>1</sup> ». En d'autres termes, Google News pourrait être accusé de concurrence déloyale. Qu'en est-il dans les faits ?

<sup>1</sup> Définition rappelée en France par la Chambre Commerciale de la Cour de Cassation dans son arrêt en date du 1<sup>er</sup> juillet 2008. Il est intéressant de noter que, dans le cadre de cet arrêt, le magazine *NewLook* a été condamné pour des actes de contrefaçon et des actes de concurrence déloyale à l'encontre du magazine *Entrevue*.

La question est complexe mais pourrait être résumée ainsi : si de fait Google News crée pour Google de la valeur économique en copiant le contenu (donc la valeur économique) produite par les médias, il génère toutefois en contrepartie du trafic – et donc de la valeur économique à travers notamment les recettes publicitaires – pour ces médias. Les médias – le cas allemand l'illustre bien – ayant été dé-référencés par Google News l'ont d'ailleurs appris au détriment de leur audience. Ainsi Google ne pourrait véritablement être accusé de parasitisme, sauf par les agences de presse qui, ne tirant pas de bénéfices financiers du trafic sur leur site (puisqu'elles fonctionnent en BtoB) mais uniquement de la vente d'abonnements à d'autres médias contre l'accès à leurs contenus, souffrent de fait du « vol » de leurs contenus par Google (**ENCADRÉ**).

### **Les médias et Google News : de nombreuses batailles juridiques**

En France, l'AFP a attaqué Google News en justice en 2005. L'AFP accusait Google d'avoir enfreint le copyright sur ses dépêches et ses photos en les diffusant sans autorisation. Plus précisément, Google Actualités affichait le titre, le chapeau et les photos d'articles publiés par différents médias, plusieurs de ces articles ayant été écrits par l'agence elle-même mais diffusés sur le site internet de tierces parties. Or l'AFP ne distribue pas son contenu gratuitement sur son site mais vend des contrats de licence pour son utilisation et les seules parties autorisées à le publier sont celles ayant payé l'abonnement. Ainsi en diffusant ce contenu, même sous forme abrégée, Google News violait le copyright et volait les contenus de l'AFP.

Avant le terme de l'affaire, Google a décidé de ne plus indexer le contenu de l'AFP. En 2007, l'AFP et Google ont conclu un accord de licence – dont le montant est demeuré secret – et les contenus de l'AFP (les articles comme les photos) sont à nouveau présents dans Google News.

De la même manière, aux États-Unis, un conflit – étudié en économie par Chiou et Tucker (2015) – a opposé Google

News et AP en 2010. En 2010, après la rupture de leurs négociations sur les contrats de licence (un premier contrat avait été signé en 2006 entre Google et AP et venait à expiration), Google News a supprimé de son site tous les contenus syndiqués par AP pendant près de deux mois. Google News n'est toutefois pas le seul agrégateur à avoir été attaqué en justice. Ainsi en 2008, AP a attaqué All Headline News qui se définissait sur son site internet comme une agence de presse et un fournisseur de contenu mondial. Cet agrégateur copiait et récrivait des articles trouvés sur Internet avant de repackager et de vendre ce contenu à des journaux et autres sites internet. La Cour de Justice a retenu les réclamations d'AP en violation de copyright et en détournement de « *hot news* » – parmi d'autres – et un accord a été trouvé entre les parties, All Headline News s'engageant à ne plus utiliser de contenu d'AP et lui versant un montant resté secret.

Ainsi on se retrouve avec Google dans une situation extrêmement paradoxale. D'une part, Google News semble ne pas porter atteinte au respect de la propriété intellectuelle du fait de l'exception de courte citation en France ou de l'exception de « *fair use* » aux États-Unis. D'autre part, Google génère du trafic pour les éditeurs – environ un quart de leur trafic proviendrait de Google selon les différentes estimations disponibles – donc un bénéfice réel. Et pourtant Google capture la plus grande part des revenus publicitaires sur Internet, alors même que l'entreprise n'investit pas dans la production de contenu, ce qui est problématique.

Si Google n'avait pas le sentiment d'une certaine responsabilité dans la crise actuelle que traversent les médias, le géant américain n'aurait sans doute d'ailleurs pas mis en place le Fonds Google – de son vrai nom Fonds pour l'innovation numérique de la presse (FINP) – en France en 2013. Doté de soixante millions d'euros sur trois ans, ce fonds « a vocation à apporter un soutien financier aux projets innovants de développement numérique portés par les éditeurs de sites d'information politique et générale ». Mais le fonds Google ne peut suffire ; ses montants ne représentent qu'une goutte d'eau dans la mer (**ENCADRÉ**).

## Le Fonds Google : une goutte d'eau

Que représentent véritablement les 60 millions d'euros dont a été doté le Fonds Google en 2013 pour trois ans en France ?

La somme peut sembler importante. Et pourtant la réalité des chiffres est toute autre.

Il ne s'agit tout d'abord « que de » 20 millions d'euros par an, c'est-à-dire moins de 10% des aides indirectes fiscales à la presse, et à peine 5% des aides directes à la presse en 2013 (le seul Fonds stratégique pour le développement de la presse s'est élevé à 33,5 millions d'euros en 2013).

De plus, cette somme de 20 millions d'euros n'est pas toujours allouée dans son intégralité. En 2013 par exemple, sur 39 dossiers déposés, 23 projets ont été sélectionnés et le montant total alloué a été de 16 millions d'euros, soit un montant moyen de 712 000 euros par projet (le Fonds soutient les projets à hauteur maximale de 60% des dépenses éligibles engagées par les éditeurs, dans la limite de 2 millions d'euros par projet). *Le Figaro* a touché 1,82 millions du Fonds Google en 2013 pour son projet Figaro.tv, *Le Monde* 1,84 millions pour « 2014 Mobilité first ». En comparaison, ces deux titres ont touché la même année plus de 16 millions d'euros d'aides à la presse. En 2014, toujours en deçà des 20 millions d'euros annoncés, le Fonds a alloué 16,1 millions d'euros à 29 projets.

En 2015, Google a annoncé le lancement du Digital News Initiative (DNI) doté lui de 150 millions d'euros pour trois ans mais à destination d'éditeurs de toute l'Europe. L'initiative a été lancée avec huit éditeurs partenaires (*Les Échos*, le *Frankfurter Allgemeine Zeitung*, *Die Zeit*, le *Financial Times*, le *Guardian*, *NRC Media*, *El Pais* et *La Stampa*). Même si ce nouveau fonds était utilisé uniquement par ces huit partenaires – ce qui ne sera pas le cas, de nouveaux éditeurs étant amenés à le rejoindre – la somme allouée ne représenterait que 6 millions d'euros par an et par éditeur, une fois encore une goutte d'eau par rapport par exemple aux aides à la presse. Et rapportée surtout aux revenus pu-

bilitaires que Google News tire de l'utilisation de contenus qu'il ne produit pas. Au niveau mondial, les recettes publicitaires de Google dépassent 67 milliards de dollars. En France, toutes les estimations s'accordent pour dire qu'elles dépassent très largement le milliard et demi d'euros. Autrement dit, les 20 millions d'euros annuels alloués au Fonds Google représentent moins de 1,5% des recettes publicitaires que Google réalise sur le marché français.

Google, par ailleurs, du fait d'un recours massif à l'optimisation fiscale, n'a payé que 6,7 millions d'euros d'impôts en France au titre de l'impôt sur les sociétés en 2015...

### *Incontournable Google ?*

L'importance prise par Google en Europe comme moteur de recherche est bien supérieure à ce qu'elle est ailleurs dans le monde<sup>1</sup>, et ce malgré l'existence d'autres moteurs de recherche comme Bing, Yahoo ! ou encore Quora<sup>2</sup>. Alors qu'en France comme en Allemagne, Google représente plus de 93% du trafic des moteurs de recherche<sup>3</sup>, aux États-Unis par exemple, d'autres moteurs de recherche font davantage concurrence au géant du web et la part de marché de Google n'est « que de » 78%.

Pourquoi une telle position dominante de Google ? En France, on a vu pourtant se développer récemment des alternatives intéressantes comme le moteur de recherche Qwant (qui totalise à peine 1,2% de parts de marché en France) qui s'est allié récemment avec Mozilla. Qwant ne trace pas les utilisateurs et ne collecte pas de données personnelles. Et ce n'est pas le seul moteur de recherche à être plus respectueux de la vie privée des utilisateurs que Google. C'est le cas par exemple de DuckDuckGo ou de Startpage qui assurent ne pas traquer les recherches effectuées. Alors pourquoi s'en tenir à Google ?

1 La Commission européenne s'est d'ailleurs attaquée en 2015 aux pratiques de Google dans le domaine des moteurs de recherche, qu'elle considère comme des abus de position dominante.

2 Même si Yahoo ! a vendu en juillet 2016 ses activités internet, dont son moteur de recherche, à Verizon (opérateur de télécommunications américain).

3 Et plus de 95% en Espagne et en Italie.

*Le droit voisin et la nécessaire réponse européenne*

L'exemple de Google – et les échecs des bras de fer espagnols puis allemands<sup>1</sup> – montre qu'en matière de propriété intellectuelle (aussi) un pays seul ne peut rien. Ainsi, si la question du droit d'auteur pour les médias doit être posée, c'est au niveau européen<sup>2</sup>.

La Commission européenne a d'ailleurs lancé au premier semestre 2016 une « consultation publique sur le rôle des éditeurs dans la chaîne de valeur du droit d'auteur et sur l'exception "Panorama"<sup>3</sup> » L'objet de cette consultation est de « recueillir les avis des parties intéressées sur l'impact que l'octroi d'un droit voisin de l'UE aux éditeurs pourrait avoir sur l'ensemble du secteur de l'édition ainsi que sur les citoyens et les industries créatives et sur la question si la nécessité (ou non) d'une intervention serait différente par rapport à la presse ou aux autres secteurs de l'édition. » Se pose également de plus en plus la question de la possible attribution d'un droit voisin aux agences de presse.

Qu'est-ce que le « droit voisin » ? C'est un droit similaire au droit d'auteur, à la différence qu'il ne récompense pas la création d'un auteur (œuvre), mais l'exécution d'une œuvre (par exemple par un chanteur) ou un investissement organisationnel ou financier (par exemple producteur) qui peut s'accompagner d'une intervention dans le processus de création<sup>4</sup>.

Il sera intéressant de connaître les résultats de cette consultation, qui ne sont malheureusement pas encore disponibles à l'heure où nous achevons ce livre. Notons cependant dès à présent plusieurs

1 Même un éditeur aussi puissant que l'éditeur allemand Axel Springer a dû renoncer à son bras de fer avec Google sur la propriété intellectuelle au bout de seulement deux semaines en octobre-novembre 2015. En effet, la chute de la fréquentation de ses sites à laquelle il a fait face après avoir interdit à Google de l'indexer a été beaucoup trop importante.

2 C'est d'ailleurs ce qu'a souligné à raison le rapport de la Commission Européenne « Vers un cadre moderne et plus européen pour le droit d'auteur » publié en 2015 : « Pour les agrégateurs d'informations, notamment, des tentatives de solutions sont nées dans certains Etats membres, mais elles risquent de morceler encore le marché unique numérique ».

3 Nous nous focalisons ici sur la question du droit voisin et ne discuterons pas de l'exception Panorama qui ne rentre pas dans le cadre de cet ouvrage. En quelques mots, l'exception Panorama est une exception au droit d'auteur par laquelle il est permis de reproduire une œuvre protégée se trouvant dans l'espace public. En France, la « liberté de panorama » est l'une des dispositions de la loi numérique.

4 Bien sûr l'octroi d'un droit voisin aux éditeurs ne doit pas se faire au détriment du droit d'auteur (et donc ici des journalistes), comme le rappelle par exemple l'arrêt rendu le 12 novembre 2015 par la Cour de Justice de l'Union européenne (CJUE) visant à préciser le régime de l'exception de reprographie (jurisprudence Reprobel).

choses. D'une part, il nous semble que l'introduction d'un droit voisin permettrait aux éditeurs de monétiser la production d'information originale, et ainsi de retrouver au moins une partie de leurs incitations à investir dans cette production et à financer le travail des journalistes. D'autre part, cette question du droit voisin ne doit pas être, selon nous, pensée uniquement du point de vue de la relation entre médias traditionnels (producteurs d'information) et agrégateurs d'information et réseaux sociaux (et notamment Facebook), mais prendre en compte également les relations entre médias traditionnels. En effet, nous l'avons montré, les médias traditionnels se copient entre eux – sans que la source originale de l'information soit toujours créditée – et il n'y a donc aucune raison pour qu'ils ne soient pas tenus juridiquement aux mêmes obligations que les autres acteurs.

La mise en place d'un droit voisin étendu aux éditeurs et aux agences de presse au niveau européen supposera enfin – pour fonctionner efficacement – la mise en place d'une (ou plusieurs) sociétés de gestion collective en charge de la collecte des redevances et de leur répartition entre les acteurs.

*Le droit d'auteur ne va-t-il pas à l'encontre d'une approche de l'information comme bien public ?*

Une limite cependant. Si l'information est un bien public, alors l'on peut souhaiter que ce bien public soit disponible au plus grand nombre. Cet argument a parfois été utilisé contre une utilisation du copyright par les médias (voir par exemple le cas sud-africain). De la même façon, il pourrait l'être contre l'introduction de murs payants. En toute logique, une fois une information produite, faire payer pour cette information revient à exclure des consommateurs qui lui accordent pourtant une valeur supérieure au coût marginal de sa diffusion (qui lui tend vers zéro).

Toutefois, pour que ce bien public soit disponible au plus grand nombre, il faut déjà dans un premier temps qu'il soit produit, autrement dit que les médias aient la capacité de financer la production d'une information originale de qualité. Cette capacité – on l'a vu – est mise à mal par le recours grandissant au copié-collé; elle l'est aussi par les moteurs de recherche et les agrégateurs qui monétisent le contenu créé par les autres acteurs (de sorte que l'on peut y voir du parasitisme au sens juridique du terme).

Mais introduire un droit voisin ne signifie nullement empêcher toute reprise de l'information et sa diffusion comme bien public. Il s'agit d'obliger ceux qui reprennent de l'information originale produite par d'autres – et en tirent un bénéfice financier d'une façon ou d'une autre – à participer au financement de cette information en payant pour l'utilisation de ce contenu.

Du point de vue des agences de presse, la mise en place d'un droit voisin permettrait de mettre fin au parasitisme et aux moteurs de recherche et aux agrégateurs de rentrer dans un cercle vertueux de financement de la production d'information, puisqu'en payant une licence aux agences – comme le font d'ailleurs les autres médias – ils contribueront à la capacité de ces médias à financer la production de contenu à travers leurs rédactions.

Du point de vue de tous les autres médias producteurs d'information originale, cela favoriserait le développement de la syndication, c'est-à-dire de la revente de contenus, qui peut prendre plusieurs formes.

## La syndication sous toutes ses formes

La syndication, c'est la revente de contenus ou *republishing*. Cette revente peut se faire dans la langue originale ou, de plus en plus, sous la forme de traductions. Le *New York Times* offre ainsi, à travers son « *news service & syndicate* », plusieurs services de traduction, en espagnol, en portugais ou encore en chinois. Pour le *New York Times*, c'est une manière de rentabiliser sa production d'information, en la vendant non seulement aux lecteurs mais également à d'autres médias<sup>1</sup>. Pour les médias qui achètent le contenu du *New York Times*, cela leur permet d'offrir à leurs lecteurs un contenu de qualité à moindre coût.

La syndication de contenus n'a bien sûr rien de nouveau – elle date du XIX<sup>e</sup> siècle – mais elle a été facilitée par le web comme nous le rappellent les initiales des flux RSS : Really Simple Syndication. Ce format de données, apparu en 2002, est en effet utilisé pour la syndication du contenu Web qu'elle facilite fortement puisque le re-

1 Même si le *New York Times* peut parfois se montrer « excessif » dans la facturation de la réutilisation de ses contenus. Le journal a ainsi demandé 1 884 dollars pour trois citations d'une longueur totale 300 mots dans le cadre d'un livre à deux chercheurs américains, Daniel Hallin et Charles Briggs, comme l'a rapporté le Nieman Lab (Shan Wang, 9 juin 2016).

cours à ce format permet aux utilisateurs de suivre automatiquement leurs sites préférés. Le flux RSS est constamment maintenu à jour afin de contenir les dernières informations à publier.

Au *Financial Times* – qui offre de nombreuses « *republishing solutions* » –, cette seule activité aurait généré 5 millions d’euros l’an dernier (selon la *Tribune*<sup>1</sup>); 2 millions d’euros en France pour un journal comme *Le Monde*, dont environ 20% à l’international. D’après le rapport annuel du *New York Times*, en 2015, la diffusion aurait généré 54% des revenus, la publicité 40% et les autres activités – qui incluent les revenus de syndication et de licences de contenu – 6%<sup>2</sup>. Mais nous ne pouvons pas isoler les revenus de la syndication et de l’octroi de licences des autres revenus, comme les revenus des archives numériques, les revenus locatifs ou encore ceux du e-commerce.

### *Paywalls ou paytags ?*

Ainsi, au-delà du développement des *paywalls*, l’introduction de « *paytags* » pourrait être une façon efficace pour un certain nombre de médias de monétiser leur production d’information originale dans le futur. C’est ce qu’a proposé Jason Fry dans un article publié sur le site du NiemanLab en 2010.

Techniquement, cela reviendrait à suivre et à tagger les contenus produits par les médias en ligne, non seulement pour s’assurer du respect des conditions d’utilisation<sup>3</sup> mais également pour monétiser la reprise de ces contenus. De nombreux logiciels ont été développés ces dernières années pour surveiller les reprises non autorisées sur Internet, en textes, en photos ainsi que de plus en plus en vidéos (**ENCADRÉ**). Mais ces logiciels ont jusqu’à présent eu principalement des fins « punitives », il n’y a qu’à penser à Hadopi. On pourrait au contraire y voir une opportunité efficace pour un certain nombre de

1 <<http://www.latribune.fr/technos-medias/medias/les-medias-rivalisent-d-innovation-pour-ne-plus-etre-uberises-491743.html>>

2 « *Syndication* » et « *licensing* » sont deux façons différentes pour un journal de revendre du contenu. D’après le *New York Times*, la syndication est « *the distribution of articles, stories, columns, images, comic strips, or other features for simultaneous publication in a number of different publishing environments and markets* ». Le « *licensing* » est défini quant à lui « *as the right to publish a branded section or magazine using the Times logo, and the same look and feel as the original* ».

3 C’est dans cette logique et dans le contexte de la doctrine des « *hot news* » qu’AP avait annoncé en 2009 le lancement d’un « *digital news registry* » dont l’objectif était de faire aux utilisateurs un « *licensing fee* ». Mais ce fut un échec.

médias – en particulier ceux dont la « marque » n'est peut-être pas assez forte pour vendre des contrats de licence comme le fait un *New York Times* – d'y monétiser « à la pièce » des articles de qualité dans lesquels ils auraient investi beaucoup de travail journalistique. Cela leur permettrait de plus de faire connaître leur marque plutôt que – comme sont contraints de le faire un grand nombre de jeunes *pure players* – de se financer en produisant du contenu en « marque blanche », autrement dit sans que leur nom ne figure.

### **Advestigo, Attributor... : les nouveaux gendarmes du web ?**

De nombreux logiciels ont été développés ces dernières années pour surveiller les reprises non autorisées sur Internet. C'est le cas par exemple de la société Advestigo ou de Copeerright Agency en France, ainsi que de Attributor aux États-Unis, spécialisées dans les technologies d'empreintes numériques ou « fingerprinting ». Le « fingerprinting » consiste à déposer et à conserver une signature sur un fichier ce qui permet ensuite de voir si ce fichier est une copie autorisée ou non, en vérifiant s'il comporte la signature, et de détecter ainsi les copies illicites sur Internet.

Autre technologie utilisée pour identifier les copies non autorisées sur Internet : le « watermarking » qui permet de rechercher dans un fichier un tatouage numérique implanté dans le fichier original.

Ces technologies ont originellement été développées par le secteur de l'industrie musicale, puis pour lutter contre le piratage de vidéos. En France, elles sont utilisées par exemple dans le cadre d'Hadopi (la fameuse Haute autorité pour la diffusion des œuvres et la protection des droits sur Internet).

Bien sûr, cela ne va pas sans poser un certain nombre de difficultés techniques, notamment en ce qui concerne le texte. Pour la photo ou la vidéo, les technologies de *fingerprinting* permettent aujourd'hui de suivre un fichier même s'il a été modifié ou dégradé, que ce soit par compression, réencodage ou encore fractionnement. Pour le texte – pour les raisons que nous avons vues tout au long de cette étude – il est d'une certaine manière plus compliqué de distinguer un fichier original puis de suivre ses reprises. La première difficulté provient de la définition même de l'« originalité » ; la simple reprise d'un communiqué de presse ne saurait par exemple être considérée comme de la production originale d'information. Ensuite, si l'on sait bien mesurer le copié-collé – ce que nous faisons ici – il est beaucoup plus dur de mettre à jour la copie dans le cas où il y a reformulation.

On ne peut pour autant ignorer aujourd'hui ces différentes technologies, tout comme d'ailleurs les systèmes de Digital Right Management (DRM) ou de gestion numérique des droits qui permettent de surveiller ce qui est fait des fichiers. Elles doivent être partie prenante du futur de la production d'information en ligne, mais cela – insistons bien – non dans une logique de punition – qui ne nous semble pas la voie à suivre – mais bien dans une logique de monétisation.

En complément des pages collectées au titre du dépôt légal<sup>1</sup>, on pourrait envisager qu'un opérateur public reçoive de manière systématique et rende accessible à la fouille de données l'ensemble des productions en ligne des médias d'information générale accompagnées de leurs métadonnées. Cette base de données pourrait également contenir les empreintes numériques de l'ensemble des articles, ainsi que des vidéos et des photos les accompagnant. Les contenus seraient directement envoyés par les médias – en contrepartie des subventions publiques – et non captés par des robots sur leurs sites web respectifs. Ils seraient accessibles aux chercheurs, voire aux médias soucieux de protéger le respect de la propriété intellectuelle de leur travail. Il nous semble en effet dommageable qu'aujourd'hui le seul accès pour les chercheurs à de telles données se fasse par des opérateurs privés. Ainsi des entreprises telles que Factiva ou Lexis-Nexis reçoivent à l'heure actuelle les articles de très nombreux organes de presse français qu'elles rendent accessibles à des coûts extrêmement élevés.

<sup>1</sup> Les sites des médias d'information sont collectés au titre du dépôt légal par la BnF pour la presse, et par l'Ina pour les médias audiovisuels.

*Le crédit avant tout*

Quelles que soient les orientations qui seront prises dans le futur, tout le monde s'accordera, nous l'espérons, pour dire que la reprise d'informations publiées par des concurrents sur Internet doit donner lieu au crédit du média repris. Nous avons montré que la réputation joue un rôle essentiel pour les médias ; créditer les médias producteurs d'information originale renforcera leur réputation et leurs incitations à en produire davantage.

Nous pensons également que l'utilisation des liens hypertextes devrait être beaucoup plus répandue. Certes la tentation est forte pour les médias – on le comprend – de vouloir garder les utilisateurs dans leur environnement. Mais la reprise – qui tient parfois à la systématisme – de contenus produits par d'autres serait sans doute beaucoup mieux acceptée par les différents acteurs si elle donnait lieu à une augmentation du trafic à travers ces liens (dans la logique des augmentations générées par les agrégateurs comme Google).

Cela permettrait peut-être aussi de rétablir une relation de confiance avec les citoyens/lecteurs trop souvent noyés sous la multiplicité de contenus similaires sur Internet. Pouvoir sourcer chacun de ces contenus – et le cas échéant retourner à la source – semble être une revendication légitime du consommateur d'information.

**Mutualiser les coûts de la production d'information**

Deux voies peuvent être envisagées pour redonner aux médias des incitations à investir dans la production d'information originale. D'une part, augmenter la monétisation de leur contenu, nous venons de le voir. D'autre part, diminuer les coûts de production de ces contenus, notamment en les mutualisant.

*Des médias « agences de presse » ?*

En syndiquant leur contenu, un certain nombre de médias font d'une certaine façon le choix du BtoB (sans oublier pour autant bien sûr leur aspect BtoC). Ils se retrouvent ainsi à leur manière dans la position d'agences de presse – même si bien sûr ils ne répondent pas à de nombreux autres aspects de ce qui définit une agence, et notam-

ment l'objectif de se concentrer sur les faits<sup>1</sup> (il est d'ailleurs intéressant de noter que ce qui fait le succès de la syndication d'un journal comme le *New York Times*, ce sont en grande partie les tribunes de ses chroniqueurs vedettes, comme Paul Krugman).

Au-delà de la syndication, nombreux sont les médias d'information générale – en France – qui ont mis en place leur propre agence, avec plus ou moins de réussite, et souvent d'ailleurs pour ne pas avoir à payer le coût jugé trop élevé de l'AFP. Ainsi en 1951, deux titres de presse quotidienne régionale, le *Provençal* et *Nord Matin*, ont lancé l'Agence Centrale de Presse (ACP) comme agence de complément pour les journaux de province. Cette agence, dont l'histoire est relativement méconnue, a conclu en 1972 un accord avec Reuters, provoquant l'ire de l'AFP qui coupa ses services à l'ACP et à Reuters qui recevait son service domestique français.

Certes, les méthodes employées par l'ACP – qui revendait donc moins chers à ses clients des contenus produits par l'AFP – sont critiquables, et si l'on peut comprendre les motivations de ses abonnés qui trouvaient là un fil moins cher que celui de l'AFP (duquel la majorité s'était désabonnée), l'on doit surtout souligner à quel point la multiplication de telles stratégies non-coopératives ne peut à terme que mettre en cause la survie même d'une agence nationale, et ce d'autant plus si cette dernière est à vocation mondiale et porte sur ses épaules une mission de service public.

Ce n'est pas pour autant que les médias doivent s'interdire de réfléchir aux conditions de mutualisation de leur production d'information. Et d'ailleurs nombreux sont ceux à l'avoir déjà fait, notamment en presse quotidienne régionale. Ainsi en 1967, *Le Dauphiné Libéré* et *Le Progrès*, qui avaient passé l'année précédente un premier accord, ont décidé de fusionner leurs rédactions avec la création de AIGLES (Agence d'Informations Générales Locales et Sportives<sup>2</sup>). Cette agence a servi régulièrement des journaux abonnés, même en dehors du groupe de presse. Dans la suite de AIGLES, est mise en place dans les années 1980 l'Agence Générale d'Informations (AGI) regroupant les rédactions du *Dauphiné Libéré*, de *Lyon Matin*, du

1 Voir la définition donnée par l'UNESCO en 1953 que nous avons citée plus haut.

2 En 1966, *Le Progrès* de Lyon et *Le Dauphiné Libéré* de Grenoble ont décidé de mettre leurs forces en commun en créant ensemble une société nouvelle qui regroupe leurs moyens techniques et toutes leurs activités : « Province n°1 ». Voir à ce sujet C. Bellanger *et al.*, *Histoire générale de la presse française*, Paris, PUF, 1976.

*Journal Rhône Alpes*, de *Loire Matin* et de *Vaucluse Matin*. L'expérience durera jusqu'à la fin de l'année 1989.

Plus récemment, le groupe EBRA (Est Bourgogne Rhône-Alpes) – propriété du Crédit mutuel qui regroupe *Vosges Matin*, *Le Journal de la Haute Marne*, *L'Est Républicain*, *L'Alsace*, *Le Bien Public*, *Le Journal de Saône et Loire*, *Le Progrès*, *Le Républicain Lorrain*, *Les Dernières Nouvelles d'Alsace*, *Le Dauphiné Libéré* ou encore *Vaucluse Matin* – a mutualisé une partie de la production de ses titres. D'une part, le groupe a mis en place en 2013 une rédaction parisienne pour traiter des informations générales. D'autre part, il incite de plus en plus chaque journal à venir « piocher » dans les contenus des autres journaux du groupe. De là à ce que réapparaisse la tentation de vendre ces contenus à d'autres titres de presse quotidienne régionale à un tarif moins élevé que celui de l'AFP...

Apparaît surtout un risque majeur, visible à la lecture des différents titres du groupe : l'uniformisation de leur contenu. Or, mutualisation ne devrait pas dire harmonisation, car sinon c'est au détriment du pluralisme et du respect de l'indépendance du travail des journalistes que celle-ci se fait. Un bureau parisien commun à l'ensemble des titres d'un groupe de presse régionale n'a de sens que s'il est en charge de la collecte de faits et qu'une liberté éditoriale est laissée à chacune des rédactions locales dans le traitement de ces faits (le groupe EBRA semble malheureusement avoir quelques difficultés à garantir la liberté éditoriale et l'indépendance des journalistes). On n'imaginerait pas que l'ensemble des titres d'un groupe collent en Une, telle quelle, une dépêche AFP.

Or la logique de la mutualisation est de plus en plus détournée à des fins d'uniformisation. Ainsi on ne peut que regretter que les sites Internet des journaux d'un même groupe (on peut par exemple penser à *La Montagne*, au *Berry Républicain*, au *Populaire du Centre* et au *Journal du Centre*) se confondent de par leur mise en page et publient simultanément les mêmes informations nationales ou internationales<sup>1</sup>. Cela va à l'encontre de la logique même des premières tentatives de mutualisation à des fins de réduction de coûts fixes mais tout en préservant le pluralisme de l'information, nous pensons par exemple à l'« American Newspaper Preservation Act » de 1970, qui

<sup>1</sup> Tout laisse d'ailleurs à penser que la publication de ces articles n'est pas gérée au niveau des titres mais par un webmaster en charge de l'ensemble de ces titres. Ce qui n'a pas été sans poser quelques problèmes aux chercheurs que nous sommes dans l'étude de la propagation de l'information.

permet à des journaux en concurrence de participer à un « *joint operating agreement* » leur permettant de partager un certain nombre d'activités commerciales tout en préservant l'indépendance de leurs activités éditoriales.

Au-delà du risque croissant d'uniformisation de l'information – que l'on retrouve d'ailleurs dans les données présentées dans ce livre, on a vu l'importance prise par le copié-collé sur Internet – le développement de synergies (c'est le terme politiquement correct) entre médias d'un même groupe se fait aujourd'hui en France de plus en plus de façon transmedia, avec la formation de grands groupes d'information multimédias (on pense en particulier à l'Altice Media Group de Patrick Drahi qui regroupe BFM TV, RMC, *Libération* ou encore *L'Express*<sup>1</sup>). Ce n'est pas le sujet de ce livre mais il ne faut pas ignorer tous les problèmes que cela pose en termes de concentration excessive de la production de l'information. Et l'on peut souhaiter que des régulations adaptées soient mises en œuvre dans le futur.

Doit-on pour autant conclure qu'il n'y a pas de place pour une mutualisation de l'information qui se ferait dans le respect de l'indépendance des journalistes et de leur liberté éditoriale ? La voie à suivre est selon nous celle des consortiums de journalistes.

### *Les consortiums internationaux de journalistes*

Les consortiums internationaux de journalistes se sont développés au cours des dernières années. L'exemple le plus connu est sans aucun doute celui du « International Consortium of Investigative Journalists » (ICIJ) ou « Consortium international des journalistes d'investigation » qui est un réseau mondial qui réunit aujourd'hui plus de 190 journalistes d'investigation dans plus de 65 pays, et qui a mis à jour récemment les Panama Papers<sup>2</sup>.

Les avantages de tels consortiums sont nombreux. D'une part, si les documents des Panama Papers ont de fait été obtenus originellement par un seul journal, le quotidien allemand *Sueddeutsche Zeitung* – auquel le lanceur d'alerte « John Doe » les a remis –, leur nombre – plus de 11 millions de fichiers ! – était de toute évidence bien trop

1 Et nous ignorons ici le développement de convergences entre médias et télécoms reposant sur la fameuse complémentarité entre tuyaux et contenus.

2 Ce groupement a été fondé en 1997 par le Center for Public Integrity à l'initiative de Charles Lewis. Autre exemple de consortium international de journalistes, le « Bureau of Investigative Journalism » basé à Londres.

important pour qu'ils puissent être traités par la rédaction d'un seul journal. Et l'on comprend pourquoi le journal allemand a fait le choix de contacter l'ICIJ et de partager ces fichiers avec les médias partenaires<sup>1</sup>. Ce n'est pas une mais plusieurs centaines d'affaires que contiennent les Panama Papers, affaires dont l'intérêt peut varier selon les pays. Il est d'ailleurs intéressant de voir comment les différents journaux impliqués ont choisi chacun à leur manière de « feuilletonner » les différentes révélations. Ainsi, *Le Monde* s'est intéressé selon les jours à la fortune offshore de l'Autorité palestinienne, à la fraude fiscale de la star du foot internationale Lionel Messi, au détournement de l'argent du pétrole par l'élite algérienne, aux diamants de Koidu en Sierra Leone, ou encore à la fortune de magnats indiens du Togo... pour ne prendre que quelques exemples.

De plus, et le cas des Panama Papers l'illustre à nouveau fort bien, certaines enquêtes, du fait même de leur nature, sont *internationales* et nécessitent donc d'agir de façon collective en dépassant les frontières. L'évasion fiscale en est un exemple typique. Ce problème peut difficilement être traité au niveau d'un seul pays, surtout lorsque les montages offshore impliquent des sociétés dont les différentes branches sont localisées aux quatre coins du globe. Or, de telles enquêtes risquent de se multiplier dans le futur du fait de la mondialisation de la société et de l'économie.

Enfin, autre avantage des consortiums de journalistes : ceux-ci permettent de limiter la censure voire l'autocensure des journalistes, ce qui est particulièrement important aujourd'hui dans le contexte accéléré de changement de mains des médias, d'apparitions de nouveaux actionnaires étrangers au secteur et de concentration croissante. En effet, on peut imaginer que des journalistes tombant sur des documents concernant leurs propres actionnaires puissent « hésiter » – ou subir des pressions – à publier les documents en question. Mais si ces documents sont entre les mains de plusieurs dizaines de journalistes travaillant pour des médias du monde entier, alors ils ne peuvent que publier les informations qui dans tous les cas seront publiées par d'autres. Ainsi, dans le cadre des Panama Papers, le journal argentin *La Nacion* a été épinglé par ses propres journalistes !

Si l'idée de l'ICIJ – et de la majorité des consortiums internationaux de journalistes à ce jour – est de se concentrer sur des problèmes qui ne s'arrêtent pas aux frontières nationales, on pourrait imaginer

<sup>1</sup> Toute comme l'avait fait d'ailleurs le journal *Le Monde* au moment du SwissLeaks.

que la même logique soit également mise en œuvre à l'échelle d'un pays. Le futur de la production d'information sur Internet passe par le développement et le renforcement du journalisme d'investigation au niveau local, national comme international.

*Mutualiser la distribution de l'information sur Internet*

Au-delà de la mutualisation de la production de l'information, il est également important aujourd'hui de penser la mutualisation de la distribution de l'information en ligne. Nous le soulignons au début de ce chapitre, le succès de *Mediapart* a poussé un certain nombre de *pure players* à fixer le prix de leur abonnement à 9€ par mois. Or, 9€ par mois, c'est la somme totale qu'un Français se dit prêt en moyenne à dépenser chaque mois pour s'informer. Ainsi, il semble urgent de repenser les modes d'abonnement aux médias en ligne.

Une partie de la solution repose peut-être sur la « vente à l'unité » : autrement dit, il ne s'agit plus de s'abonner à un seul média, mais de passer par une plateforme regroupant les sites internet de plusieurs médias et vendant à l'unité les articles de ces médias à un prix extrêmement faible (quelques centimes). C'est ce que propose la plateforme Blendle, start-up hollandaise surnommée le iTunes de la presse<sup>1</sup> qui, aux États-Unis a déjà passé un partenariat avec le *New York Times*, le *Washington Post* ou encore le *Wall Street Journal*, et qui offre également à lire les contenus de *The Economist*<sup>2</sup>. Ainsi, les fameux 9€ par mois seraient répartis entre plusieurs médias, et l'on peut penser que des lecteurs hésitant à dépenser cette somme pour un seul média franchiraient le pas s'ils pouvaient ainsi avoir accès à une multiplicité de contenus. Une telle plateforme permettrait également de toucher tous ceux qui n'ont pas 9 mais 1 ou 2 euros seulement à dépenser par mois pour l'information et qui malgré tout pourraient ainsi acheter des articles.

1 Mais qui ressemble davantage à Spotify.

2 La plateforme a par ailleurs d'ores-et-déjà convaincu une grande partie de la presse allemande. Quid de la France ? Alors que *Libération* avait donné son accord de principe fin 2015, les autres éditeurs se montrent plus frileux. Au-delà des problèmes que pourrait poser une mise en concurrence de cette plateforme avec les sites internet des journaux – mais les expériences néerlandaises et allemandes nous montrent que cela n'a pas affecté négativement le trafic des sites, les utilisateurs n'étant tout simplement pas les mêmes – le véritable enjeu est celui de la marque. On l'a vu en effet tout au long de ce livre, la réputation est devenue une dimension essentielle dans un monde où l'information se propage plus vite que jamais. Or les éditeurs peuvent craindre d'affaiblir leur marque en la « mélangeant » sur une même plateforme avec d'autres marques.

Un autre avantage de la vente à l'unité d'articles est qu'elle peut être vue comme une source supplémentaire de revenus pour les médias si elle pousse des utilisateurs qui ne se seraient jamais abonnés à payer pour lire du contenu. L'utilisateur moyen de Blendle est beaucoup plus jeune que l'abonné moyen des journaux présents sur la plateforme, même si l'on ne considère que les abonnés numériques. La vente à l'unité est-elle appelée à se développer davantage dans le futur ? Il faut mettre dans la balance, en vis-à-vis du développement de Blendle, l'échec de NewsStand, le kiosque numérique d'Apple lancé en 2011 et qui permettait lui aussi aux éditeurs de vendre non pas des articles mais des journaux numériques à l'unité<sup>1</sup>.

En France, « La Presse Libre » s'est récemment lancée avec le slogan : « Un seul paiement, un seul site : l'abonnement simple et rapide à la presse en ligne<sup>2</sup> ». Ce site permet de s'abonner à plusieurs médias en ligne – à l'heure où nous terminons ce livre Next INpact, @rrêt sur images, Hors-série et Les Jours, Alternatives Économiques est également annoncé –, en un seul paiement et à un tarif préférentiel. Quel avantage de « La Presse Libre » face aux mastodontes évoqués plus haut ? Le site est géré directement par des éditeurs français<sup>3</sup>. Il ne s'agit ici en aucune manière de faire du nationalisme déplacé. Mais il est important de noter l'importance croissante prise par les fameux GAFKA – d'Instant Articles de Facebook au News d'Apple – comme acteurs majeurs de la distribution de contenu. Ces acteurs, et c'est effrayant, se placent chaque jour un peu plus comme intermédiaires de distribution entre les médias et leurs lecteurs, ce qui est une menace réelle pour la liberté éditoriale des médias à terme. On pourra d'ailleurs lire à ce sujet les réflexions extrêmement intéressantes d'Emily Bell, directrice du Tow Center for Digital Journalism, sur l'importance croissante des réseaux sociaux, devenus selon elle plus puissants que les médias<sup>4</sup>.

Ainsi, au-delà des GAFKA, on peut espérer que si une plateforme de ventes à l'unité devait se développer en France, ce ne soit pas Blendle, mais un Blendle français. D'autant que les éditeurs français ont largement les moyens de mettre en œuvre une telle plateforme. La distribution papier de la presse française est depuis la loi

1 NewsStand a d'ailleurs été remplacé en 2015 par une nouvelle plateforme, Apple News, concurrente directe de l'Instant Articles de Facebook.

2 <<https://beta.lapresselibre.fr>>

3 Et de plus, soulignons-le car cela a son importance, par des éditeurs français *indépendants*.

4 Bell, E. (2016). « Facebook is eating the world », *Columbia Journalism Review*, 7 mars 2016.

Bichet de 1947 un système coopératif mutualisant les coûts tout en s'appuyant sur un opérateur privé. Les administrateurs de Presstalis regroupent les représentants de la Coopérative des magazines d'une part et de la Coopérative des quotidiens de l'autre. Pourquoi ne pas entamer autour de la table – en incluant bien sûr les *pure players* dans la discussion – une réflexion sur la mutualisation de la distribution de l'information en ligne ? Certes Presstalis souffre d'une image datée et l'évoquer comme une partie de la solution dans une réflexion sur l'avenir des médias sur Internet fait parfois sourire. Mais Presstalis sait aussi innover en ligne, comme en a d'ailleurs témoigné le lancement de Zeens, application mobile permettant de trouver son journal en kiosque.

### **Un soutien public transmedia à la production d'information**

Les médias d'information générale produisant un bien public, il est nécessaire de penser les conditions d'un soutien public à cette production. Évaluer l'efficacité des aides de l'État aux médias dépasse très largement l'objectif de ce livre, de même que discuter de l'ensemble des réformes qui devraient y être apportées<sup>1</sup>. Il est important néanmoins de souligner que le financement public des médias doit être approché dans sa globalité, c'est-à-dire en prenant en compte à la fois les aides de l'État à la presse (et aux sites d'information en ligne) et le financement de l'audiovisuel public.

Nous avons discuté plus haut de la question du copyright pour les médias. Plutôt que de renforcer le droit d'auteur pour offrir plus de protection aux médias, certains proposent que les médias abandonnent le droit d'auteur au profit de subventions publiques. L'État finance d'ailleurs dans de nombreux pays la production d'informations originales à travers le financement de l'audiovisuel public qui se décline aujourd'hui également sur Internet.

<sup>1</sup> On pourra se reporter sur ce sujet à Cagé, J. et Fize, E. (2016). « The Effectiveness of Public Support to the Press. Evidence from France », document de travail Sciences Po Paris. Le livre *Sauver les médias* (*op. cit.*) discute également de « l'illusion des médias sous assistance » et du vrai poids des aides à la presse.

*Un système commun de l'information ?*

Dans son « projet pour une presse libre<sup>1</sup> », le journaliste Pierre Rimbert pense la production d'une information de qualité soustraite à la loi du marché comme aux pressions du pouvoir. Il propose la création d'un « service mutualisé d'infrastructures de production et de distribution de l'information » qui intégrerait – entre autres – les imprimeries, le papier, les messageries, les kiosques, les locaux, les serveurs, les outils de stockage et de diffusion, les moyens de recherche et de développement, les services administratifs, comptables, juridiques, commerciaux... Le plus simple est peut-être de souligner ce qui n'y serait pas intégré : les seuls journalistes.

Ce projet est extrêmement intéressant. Nous soulignons la nécessité de plus en plus forte pour les médias de mutualiser une partie de leurs coûts de production tout en préservant l'indépendance et la liberté éditoriale des journalistes. C'est ce que permettrait, d'une certaine manière, ce système commun. De plus, cette solution innovante a le mérite de rappeler que le « bon » modèle pour la production de ce bien public qu'est l'information est sans doute le modèle non lucratif. Pour profiter du Service commun proposé par Pierre Rimbert, la presse d'information d'intérêt général devrait en effet adopter le statut d'entreprise à but non lucratif.

Soulignons toutefois ce que nous semblent être les limites de cette solution. La première, ce sont les conditions concrètes de la mise en place d'un tel niveau d'intégration qui relève presque de l'utopie. Pierre Rimbert passe peut-être trop rapidement sur la question du financement d'un tel service commun. Il propose de faire table rase de l'intégralité des aides à la presse en faisant le pari d'un déversement suffisant des recettes publicitaires de la presse d'information vers la presse récréative, ce qui semble irréaliste.

De plus, Pierre Rimbert tend à ignorer exagérément l'existant. Il ne fait référence à l'AFP qu'au détour d'une phrase, en soulignant que le Service commun engloberait « les infrastructures de l'AFP ». Mais l'AFP vaut bien plus par ses journalistes (qui sont exclus du Service commun proposé par Pierre Rimbert) que par ses infrastructures. Or, comme nous l'avons rappelé, l'Europe a reconnu en 2015 la mission d'intérêt général de l'AFP et, depuis cette date, la contribution financière de l'État en faveur de l'AFP se décompose en un

<sup>1</sup> *Le Monde diplomatique*, 729, décembre 2014.

abonnement aux services de l'AFP d'une part, et en une subvention pour compensation du coût net des missions d'intérêt général d'autre part. Pierre Rimbert ne l'explique pas entièrement mais, de toute évidence, ses calculs visant à supprimer l'intégralité des aides à la presse suppriment également cette subvention. Pourquoi faire table rase d'un système visant à produire de manière non lucrative (c'est le cas de l'AFP) une information de qualité qui profite à l'ensemble des médias d'information politique et générale ? De plus, Pierre Rimbert se fait l'avocat de la mise en place d'un Service commun de distribution sans même mentionner le fait qu'en France, depuis la loi Bichet de 1947 (qui est une spécificité française et qui explique la richesse du nombre de titres qui sont aujourd'hui publiés en France par rapport à de nombreux autres pays), « le groupage et la distribution de plusieurs journaux et publications périodiques ne peuvent être assurées que par des sociétés coopératives de messageries de presse ». Sans même évoquer également que si Presstalis existe déjà, tout le monde s'accorde aujourd'hui sur la nécessité de le réformer. Pourquoi ne pas penser une réforme – nécessaire – de la distribution coopérative de la presse en France plutôt que de proposer un nouveau Service commun comme si rien de tout cela n'existait ?

*Une solution alternative : la société de média à but non lucratif*

Pierre Rimbert a raison d'affirmer que les médias français (au-delà de la seule presse) ont besoin d'un modèle radicalement différent. Dans *Sauver les médias*, l'un des auteurs de ce livre a d'ailleurs proposé un nouveau modèle économique et juridique pour les médias : la société de médias à but non lucratif, à l'intersection entre la fondation et la société par actions, qui vise lui aussi à soustraire les médias à la seule loi du marché. La priorité de ce modèle est de donner davantage de pouvoir de décision aux journalistes d'une part et aux lecteurs de l'autre, afin d'assurer la production d'une information de qualité en toute indépendance.

Quels sont les avantages de la fondation ? Premièrement, les « investisseurs » ne peuvent récupérer leur investissement en capital, ce qui rend celui-ci pérenne et limite la crainte des journalistes de déplaire à leur actionnaire qui ne peut reprendre ses billes, et donc l'autocensure, quand ce n'est pas la censure elle-même. Deuxièmement, les éventuels profits sont entièrement réinvestis, ce qui permet d'investir pour l'avenir plutôt que d'avoir une recherche de profits

à court terme. Enfin, dans un modèle de fondation, les apports en capital sont défiscalisés. Il y a donc une vraie incitation à investir. Jusqu'ici cette défiscalisation a été réservée aux actions caritatives dans le domaine de la santé, de l'éducation, des arts, etc., et les médias en ont été très largement exclus, en dehors de quelques cas très particuliers. Il est temps de reconnaître que la production d'informations utiles au débat démocratique est une œuvre d'utilité publique, autant que toutes les autres.

Quelle est la limite du modèle de fondation? Aujourd'hui, lorsqu'une fondation est créée, généralement ce sont ceux qui réalisent le premier investissement en capital qui tiennent la plume au moment où les statuts sont écrits et qui composent le premier conseil d'administration. Et, bien souvent, s'il y a une chose qu'ils ne partagent pas, c'est le pouvoir. Le conseil d'administration va avoir tendance à se coopter de génération en génération. Donc si les donateurs ont renoncé à récupérer leur capital, s'ils ont résolu le problème des successions, s'ils réinvestissent les dividendes, ils gardent toutefois tout le pouvoir de décision. C'est le cas par exemple du groupe Bertelsmann en Allemagne ou de ProPublica aux États-Unis. Et même quand la composition du conseil administration se diversifie, on observe généralement une certaine inertie, un manque de dynamisme et de renouvellement, qui est pourtant indispensable dans le secteur des médias (davantage peut-être que pour un musée ou pour une université).

Avec la société de média à but non lucratif, le partage du pouvoir est beaucoup plus équitable car les petits donateurs, qui aujourd'hui dans un modèle de crowdfunding ou financement participatif ne seraient que des donateurs, deviennent de petits actionnaires. Chaque apport en capital donne lieu à des droits de vote. Et la société de média à but non lucratif va plus loin en majorant les droits de vote des petits actionnaires plus que proportionnellement à leur apport en capital. Ainsi pour les petits actionnaires, disons en dessous de 10%, les droits de vote augmentent plus vite que l'apport en capital et, à l'inverse, les plus gros actionnaires, au-dessus de 10%, ne voient leur droit de vote n'augmenter que pour un tiers avec leur apport en capital. Cela pourrait se faire facilement avec l'utilisation de deux types d'action, d'une part des actions ordinaires à droit de vote unique, et de l'autre des actions préférentielles à droits de vote multiples.

Il s'agit ainsi de favoriser le pouvoir des plus petits actionnaires – qu'ils soient journalistes, salariés ou lecteurs – et de limiter la

toute-puissance des plus gros aux ressources infinies. L'idée est de repenser le partage entre le capital et le pouvoir à l'intérieur de la société de média. Nous estimons que les médias ne sont pas des entreprises comme les autres. Ils ne peuvent pas être juste possédés et gérés par des milliardaires aux poches sans fond. À ce sujet, la crise entre les journalistes d'I-Télé et l'actionnaire majoritaire du groupe, le puissant Vincent Bolloré, est révélatrice de la tentation, pour des acteurs extérieurs au secteur des médias, de faire des canaux qu'ils acquièrent des outils d'influence et d'instrumentaliser les rédactions pour les mettre au service de leurs intérêts. Ce statut hybride permettrait ainsi d'œuvrer pour des médias *indépendants* des actionnaires extérieurs, des annonceurs et des pouvoirs publics, mais *dépendants* de leurs lecteurs, de leurs salariés et des internautes. Il s'agit d'un modèle économique adapté à la révolution numérique et aux enjeux du XXI<sup>e</sup> siècle, qui tente de répondre conjointement à la question de la démocratie et du renouvellement.

Il s'agit également de reprendre l'ambition d'une presse indépendante et démocratique qui avait été exprimée en France à la Libération (avec les sociétés coopératives, puis le développement dans les années 1950 des sociétés de rédactions et de lecteurs) en apportant des outils et une flexibilité adaptés au monde actuel : des apports en capital différents doivent pouvoir donner lieu à des droits de vote différents (ce qui est en principe interdit dans les coopératives), pourvu que le pouvoir ne soit pas confisqué par les plus gros actionnaires.



## CONCLUSION

Pierre Bourdieu notait dans son essai *Sur la télévision* que la poursuite du scoop impose une contrainte terrible : « Pour être le premier à voir et à faire voir quelque chose, on est prêt à peu près à n'importe quoi, et comme on se copie mutuellement en vue de devancer les autres, de faire avant les autres, ou de faire autrement que les autres, on finit par faire tous la même chose ; la recherche de l'exclusivité, qui, ailleurs, dans d'autres champs, produit l'originalité, la singularité, aboutit ici à l'uniformisation et à la banalisation<sup>1</sup>. »

L'uniformisation et la banalisation. Les mots sont durs, le jugement violent. L'argument pourtant porte et interroge plus encore aujourd'hui, alors que quelques secondes suffisent à la propagation d'une nouvelle. Ballotés entre les mutations technologiques, la course à l'hyper-réactivité, l'appétit des infomédiaires et la gratuité comme maxime de consommation, les acteurs médiatiques peinent à trouver le bon modèle pour monétiser leur activité. Et ce, malgré un foisonnement d'innovations et d'expérimentations, et les prises de risque de nombreux journalistes dans un contexte de frilosité actionnariale.

Alors, l'information à tout prix ? Il paraît nécessaire de questionner d'une part le fonctionnement des médias et plus spécifiquement leurs structures et les contraintes économiques qui pèsent sur la production d'information. Qui finance la production d'information et peut-on encore la monétiser ? Qu'en est-il de la réalité des données ?

L'étude quantitative de la production d'information en ligne par les médias français sur une année nous a permis de mettre à jour un certain nombre de faits indispensables à une réforme en profondeur d'un secteur médiatique en quête de solutions.

Tout d'abord, les journalistes sont indispensables à la production d'une information originale. Si rien n'est fait et que la diminution

---

1 Bourdieu, P. (1996). *Sur la télévision*. Paris, Liber-Raison d'agir.

de la taille des rédactions continue au rythme actuel alors, malgré le sentiment diffus d'un foisonnement de l'information dans un univers médiatique où l'on se perd de plus en plus, moins de *faits* seront révélés demain qu'aujourd'hui. Ce ne sera sans doute pas sans conséquences sur le choix des personnels politiques amenés à nous gouverner. Toute réforme des médias doit mettre en son sein la question de la taille des rédactions, de leur financement et de leur indépendance.

De plus, du fait de la propagation extrêmement rapide de l'information et de la prévalence du copié-collé, les bénéfices « instantanés » en termes d'audience de la production d'information par les médias – malgré le coût élevé de cette production – sont relativement faibles. En revanche, le cumul dans le temps de la production d'informations originales, résultat sur le long terme du travail de ses journalistes, permet d'asseoir la réputation d'un média. Cette réputation est un facteur décisif pour expliciter les audiences. Les médias ont donc plus que jamais une incitation à produire une information originale de qualité au jour le jour : leur réputation. Dans ce cadre, leurs efforts doivent être clairement établis aux yeux des lecteurs, d'où la nécessité d'être dans le futur beaucoup plus attentif que nous ne le sommes aujourd'hui au respect des références et des crédits.

L'importance de la réputation tend d'ailleurs à faire émerger de nouvelles pistes en termes de monétisation de l'information. Le besoin de rentabilité économique à court terme pourrait bien étayer le développement du droit d'auteur ou de droits voisins au domaine des médias. La syndication de contenus commence d'ores et déjà à devenir une source de revenu pour des médias à forte réputation. Quant à la syndication de la diffusion, si elle ne touche pour l'instant que les nouveaux venus en ligne, elle permet tout en contournant la voracité des GAFAs, d'expérimenter une nouvelle relation au public, à savoir l'article à la demande dans une offre multiple, ou un choix informationnel adapté à un usage instantané, borné par le revenu ou le temps disponible.

Car il nous faut également questionner notre relation de citoyens consommateurs – et pourtant si souvent mauvais payeurs – à l'information. Quelle information souhaitons-nous collectivement et à quel prix ? Secteur d'activité à part dans le tissu économique, média et information tiennent une place essentielle pour assurer la vitalité de nos démocraties. On est prompt à souligner leurs manquements, encore faudrait-il rappeler leur utilité. On ne peut éluder cette question. L'objectif à poursuivre est bien d'assurer une information la

plus complète possible des citoyens et de garantir l'expression de la diversité des points de vue.

À l'heure où nous écrivons, le Brexit est voté, Donald Trump est devenu président, François Fillon a remporté les primaires de la droite. Les faits font des pieds de nez à tous les pronostics et à la majorité des opinions développées dans les médias. Tentation du bon mot? Opinion maternante? Infotainment? Reflet de l'injonction de l'élite? Rupture insupportable entre paroles et réalité? Manque de moyens? Manque d'intérêt? Quels contenus pour quels effets? Quid du rôle des médias? Avec un pourcentage de confiance et de crédibilité en baisse continue, la relation entre les médias et les citoyens est plus fragile que jamais... au risque de modifier les équilibres sociétaux. Le terme – aberrant! – d'« ère post-vérité » émerge... à l'heure où, paradoxalement, le fact checking tend à s'imposer comme une ressource informationnelle innovante et une composante essentielle pour tenter de rétablir cette relation entre médias et citoyens.

Il n'y a pas de solution clefs en main. Mais une partie de la réponse à cette crise inédite de confiance repose dans le rétablissement de liens réels entre les journalistes et les citoyens. Cela peut passer par des citoyens à nouveau – n'oublions pas que cela a été le cas historiquement pour certains journaux de référence! – actionnaires de leurs médias, ne se sentant pas considérés comme du temps de cerveau disponible offert par des actionnaires cupides à des publicitaires, mais partie prenante de l'aventure de l'information. Cela peut également passer par un débat de fond sur les modes de recrutement et de formation des journalistes aujourd'hui. Journalistes parfois précarisés mais souvent aussi formatés, dont il faut à tout prix sauvegarder l'indépendance et la liberté de penser.

Oui, le paysage médiatique est constamment réactif et innovant comme le prouve l'émergence de nouveaux champs et acteurs. Et pourtant, inexorablement, le nombre de journalistes baisse dans les rédactions. Derrière l'abondance du contenu informationnel disponible en ligne, qualifiée d'infobésité, se dessinent de plus en plus un appauvrissement rédactionnel et une baisse globale de l'originalité des articles publiés. Les décideurs, que ce soit les actionnaires ou les pouvoirs publics, ont une responsabilité importante dans cet état de fait. Probablement pour des raisons différentes et parce que les objectifs des uns et des autres ne sont pas, tant s'en faut, en phase, les solutions satisfaisantes tardent à émerger, en France comme ailleurs. Faire en sorte que le cadre économique, réglementaire et que

les outils de régulation et les aides à la presse soient en adéquation avec le rôle des médias dans une société elle-même en mutation, passe nécessairement par une connaissance approfondie du monde médiatique et une évolution des pratiques et des mentalités. Ce livre se veut avant tout un appel au débat, à la discussion, et à la collaboration. Espérons que nombreux soient les journalistes et les citoyens qui y répondront.

## TABLE DES MATIÈRES

### INTRODUCTION .....9

<i>Une analyse inédite des contenus produits sur Internet.....</i>	<i>10</i>
<i>Gratuité.....</i>	<i>11</i>
<i>Copié-collé.....</i>	<i>12</i>
<i>Fragilité ?.....</i>	<i>13</i>
<i>Citations et réputation.....</i>	<i>17</i>
<i>Pour une régulation des médias adaptée à Internet.....</i>	<i>19</i>
<i>Limites de l'analyse.....</i>	<i>20</i>

### CHAPITRE 1. L'INFORMATION SUR INTERNET :

#### UNE APPROCHE TRANSMEDIA.....23

Une base de données unique.....	23
Une approche transmedia.....	27
Une information... en continu ?.....	31
En France en 2013 : 25 000 événements.....	34
<i>Un algorithme de détection d'événements.....</i>	<i>36</i>
<i>Les événements : quelques éléments de description.....</i>	<i>39</i>
<i>La production d'information hors événements.....</i>	<i>42</i>
La propagation de l'information vue à travers le prisme des événements.....	43
<i>Une vitesse de propagation extrêmement élevée.....</i>	<i>44</i>
<i>La réactivité comme seul objectif ?.....</i>	<i>48</i>
<i>L'importance du copié-collé.....</i>	<i>49</i>
<i>Le risque croissant de l'homogénéisation.....</i>	<i>54</i>

## CHAPITRE 2. LA PROPAGATION DE L'INFORMATION

### SUR INTERNET : ILLUSTRATIONS .....57

<i>Le Monde</i> , Snowden et les écoutes de la NSA.....	57
<i>Une illustration du découpage en événements</i> .....	66
<i>Mediapart</i> et l'enquête judiciaire sur le micro-parti de Marine Le Pen.....	68
L'AFP annonce la candidature de François Fillon à la primaire.....	70

## CHAPITRE 3. LES DÉTERMINANTS ÉCONOMIQUES

### DE LA PRODUCTION D'INFORMATION .....73

Le coût de l'information : les journalistes.....	73
<i>La taille des rédactions</i> .....	74
Un détour par les robots.....	76
<i>Le traitement automatique du langage naturel</i> .....	76
<i>Avantages et limites des robots</i> .....	77
Production d'information et taille des rédactions : éléments chiffrés.....	79
<i>Une première approche</i> .....	81
<i>Estimations économétriques</i> .....	88
<i>Prochaines étapes : investigation et réseaux sociaux</i> .....	92

## CHAPITRE 4. LES AGENCES DE PRESSE :

### DES ACTEURS À PART .....95

Une production d'information « à part ».....	96
La plus grosse rédaction de France.....	98
<i>Les autres facteurs de la réussite de l'AFP</i> .....	98
L'AFP : rappels historiques.....	99
Un modèle BtoB.....	102
<i>L'AFP, le seul acteur au modèle économique approprié ?</i> .....	103
<i>Le développement du copié-collé... à travers le filtre de l'AFP ?</i> .....	104
Une mission d'intérêt général.....	104

## CHAPITRE 5. LA MONÉTISATION DE L'INFORMATION ..... 107

À qui profite la production d'information originale ?.....	108
<i>Les variations de l'audience agrégée</i> .....	108

<i>Les déterminants de la part d'audience de chaque média d'information</i> .....	111
<i>Les utilisateurs récompensent-ils les médias pour leur réactivité ?</i> .....	114
<i>Des effets non seulement faibles mais également non persistants</i> .....	115
Comment expliquer la faiblesse des variations d'audience ?.....	116
<i>Quelle(s) porte(s) d'entrée pour la consommation d'information ?</i> .....	116
Les recettes publicitaires en ligne, une source de revenu pour les médias d'information ?.....	117

## **CHAPITRE 6. L'IMPORTANCE DES CITATIONS ET DE LA RÉPUTATION..... 119**

Références et réputation .....	119
<i>Un algorithme de détection de citations</i> .....	120
<i>Les médias les plus copiés reçoivent-ils davantage de citations ?</i> .....	123
La production d'information originale explique-t-elle l'audience moyenne ?	125
Les variations des ventes « hors ligne ».....	126
Un nouvel enjeu pour les médias : garder les utilisateurs dans leur environnement.....	127

## **CHAPITRE 7. QUELLES SOLUTIONS ? ..... 129**

<i>Un paysage médiatique en bouillonnement</i> . . . .....	129
<i>... qui témoigne tout autant d'une grande vitalité que d'une grande fragilité</i> .....	130
Les murs payants : une solution déjà mise à mal ?.....	133
<i>Le Sun, crépuscule des murs payants ?</i> .....	133
<i>Instant Articles : la fin des paywalls ?</i> .....	134
<i>Avant le mur payant, penser au copyright ?</i> .....	136
Avantages et limites du copyright.....	138
<i>L'absence de protection des faits</i> .....	139
<i>L'exception de courte citation</i> .....	142
<i>Le droit d'auteur et les agrégateurs</i> .....	143
<i>Incontournable Google ?</i> .....	147
<i>Le droit voisin et la nécessaire réponse européenne</i> .....	148
<i>Le droit d'auteur ne va-t-il pas à l'encontre d'une approche de l'information comme bien public ?</i> .....	149
La syndication sous toutes ses formes.....	150
<i>Paywalls ou paytags ?</i> .....	151

<i>Le crédit avant tout.....</i>	<i>154</i>
Mutualiser les coûts de la production d'information.....	154
<i>Des médias « agences de presse » ?.....</i>	<i>154</i>
<i>Les consortiums internationaux de journalistes.....</i>	<i>157</i>
<i>Mutualiser la distribution de l'information sur Internet.....</i>	<i>159</i>
Un soutien public transmedia à la production d'information.....	161
<i>Un système commun de l'information ?.....</i>	<i>161</i>
<i>Une solution alternative : la société de média à but non lucratif.....</i>	<i>163</i>
<b>CONCLUSION.....</b>	<b>167</b>



Composition : Le Bord de L'eau éditions  
www.editionsbdl.com

Cet ouvrage a été achevé d'imprimer en février 2017  
pour le compte de INA Éditions  
par XXXXXXXXXXXXXXXXXXXX

Dépôt légal : mars 2017

*Imprimé en France*



# L'information à tout prix

Le développement de l'information sur Internet conduit-il au règne du copié-collé, au détriment de l'information de qualité? Peut-on inventer de nouveaux modèles économiques pour les médias permettant de tirer le meilleur parti des nouvelles technologies?

Cet ouvrage apporte de nouvelles réponses à ces questions, en conjuguant les outils du *big data*, du *machine learning* et de l'économie. Il se fonde sur la construction et l'analyse d'une base de données unique : l'intégralité du contenu produit en ligne par les médias d'information en France sur une année (2013), qu'il s'agisse de la presse écrite, de la télévision, de la radio, des *pure internet players* ou encore de l'AFP.

En appliquant leur algorithme de détection de copie, les auteurs montrent que, dans le cas des actualités « chaudes », 64 % de l'information publiée en ligne correspond à du copié-collé pur et simple, créant un niveau d'homogénéité insoupçonné. Ce recours croissant au copié-collé, combiné à une vitesse de propagation extrêmement élevée de l'information en ligne, risque de tuer à terme les incitations des médias à produire de l'information originale.

Au-delà de ces constats, les auteurs proposent des solutions économiques et juridiques pour lutter contre la dévalorisation du travail journalistique. Il en va de notre démocratie : Internet va-t-il demeurer une chance unique de mieux informer les citoyens ou peut-il demain se transformer en tombeau de l'information?

Julia CAGÉ, normalienne, est Professeur d'Économie à Sciences Po Paris et chercheur affiliée au Centre for Economic Policy Research. Titulaire d'un doctorat d'économie de l'université d'Harvard, elle est l'auteur de *Sauver les médias. Capitalisme, financement participatif et démocratie*, Paris, Le Seuil, 2015.

Nicolas HERVÉ est Ingénieur de Recherche à l'Ina, titulaire d'un doctorat d'informatique de l'université Paris-Sud Orsay. Ses travaux portent sur l'analyse automatique et l'indexation de contenus multimédia.

Marie-Luce VIAUD est Chercheur à l'Ina, titulaire d'un doctorat d'informatique de l'université Paris-Sud Orsay. Elle est spécialisée en fouille et visualisation de données. Elle coordonne avec Nicolas HERVÉ la plateforme de recherche OTMedia.

17€

