

# Ce que font les données au journalisme

## *Un point de vue du design*

< Florian Harmand <sup>1</sup> >

1. MICA (Médiation Information Communication Arts), Université Bordeaux Montaigne  
10 Esplanade des Antilles, 33607 Pessac, France  
florian.harmand@gmail.com

DOI : 10.25965/interfaces-numeriques.4435

### < RÉSUMÉ >

Cet article propose d'étudier, au prisme du design, comment les données transforment la pratique du journalisme. Bien qu'elle ne constitue pas une rupture majeure dans le métier de journaliste, l'introduction de données dans la fabrique de l'information n'est pas neutre pour la profession, au sens où elle modifie le cadre de la production de l'information (Gitlin, 1980 ; Hackett, 1984). En favorisant d'un côté une nouvelle capacité d'enquête du journaliste et de l'autre une nouvelle rhétorique, le journalisme de données exacerbe un dilemme aussi vieux que le journalisme lui-même : l'objectivité de l'information face à la subjectivité de ceux qui la façonnent.

### < ABSTRACT >

Through the prism of design, this article proposes to study how data transforms the practice of journalism. Although it does not constitute a major break in journalism, the introduction of data into the fabric of information is not neutral for the profession: it modifies the information production frame (Gitlin, 1980; Hackett, 1984). By favorizing on the one hand a new investigative capacity of the journalist and on the other a new rhetoric, data journalism exacerbates an old journalism's dilemma: the objectivity of information faced with the subjectivity of those who shape it.

### < MOTS-CLÉS >

journalisme, données, information, design, objectivité, technique, rhétorique

**< KEYWORDS >**

journalism, data, information, design, objectivity, technology, rhetoric

---

**1. Introduction**

Questionner « ce que les données font au journalisme » revient à interroger la façon dont les données et leurs techniques afférentes modifient l'exercice de la profession. Cette réflexion croise des données historiques, des enseignements du design et des projets représentatifs du journalisme de données. Elle s'attache à définir le cadre de la production de l'information investi par la notion d'objectivité journalistique et à expliquer depuis le design comment les données modifient ce cadre de production.

Dans un premier temps, quelques détours historiques relèveront des éléments constitutifs de l'objectivité à travers les époques en les reliant aux conditions de production de l'information (Feyel, 2006). Une fois ce contexte dressé, l'objectivité sera mise en tension sous deux modes : celui de la distorsion et celui de la construction.

Dans un second temps, les disciplines du journalisme et du design dialogueront autour du travail de la donnée pour en cerner l'influence sur les manières de construire et de communiquer l'information. Ce dialogue interdisciplinaire comportera deux segments de réflexion : la donnée comme nouvel outillage des professionnels et comme nouvelle rhétorique. La progression sera didactique : du cadrage disciplinaire et terminologique jusqu'à l'explicitation des potentialités et biais que la donnée apporte au journalisme.

L'article s'achèvera en assumant que le journalisme de données s'inscrit dans une continuité de l'évolution du métier de journaliste, l'occasion de souligner les perspectives éthiques, méthodologiques et téléologiques que la sous-discipline offre à la méta-discipline.

## 2. Le cadre de la production

### 2.1. *Information partisane, information objective*

Le design s'intéresse à la façon dont le processus et le contexte conditionnent le produit. Ce postulat paraît transposable au journalisme, au sens où l'information est déterminée par son cadre de production. Pour éclairer cette interdépendance, et avant de l'aborder dans son milieu contemporain et numérique, on peut en repérer les gènes dans la brève et intense histoire du journalisme. En effet, de l'encre a coulé sous les presses de la première *Gazette* à nos jours, et ce laps de temps contient autant de manières de produire de l'information que de « conceptions » de l'information<sup>1</sup>.

Prototype du journal, *La Gazette* de Théophraste Renaudot (1631) émerge dans un contexte bien différent de notre presse contemporaine. Privilégiée par le Roi et son gouvernement, monopolisant la diffusion de l'information, *La Gazette* est empêchée d'« ambitions mercantiles » (Feyel, 2003). Malgré ce carcan, Renaudot esquisse un *credo* journalistique : informer la population, s'ériger d'utilité publique, divertir le lecteur, recouper des sources diverses ou encore empêcher les « faux bruits » (ancêtre des *fake-news* ?). La fiabilité de cette information a comme corollaire ses nombreux cas d'usages : aiguiller les commerçants vers les marchés et les soldats vers les guerres ou soutenir les correspondances épistolaires (Renaudot, 1631, 5).

Au XIX<sup>e</sup> siècle, la presse est plurielle et les journaux ouvertement politisés animent le débat public. Mais en 1826, un gouvernement royaliste menace sa pérennité en augmentant le droit de timbre sur les publications et les tarifs postaux. Les responsables de journaux d'opinion sont contraints de repenser leur format de façon à incorporer des annonces pour équilibrer les finances (Feyel, 2003). Le travail éditorial se double d'une ingénierie financière, le journal commence à être pensé

---

<sup>1</sup> Les citations de *La Gazette*, du *National* et de *La Presse* sont issues de la plateforme Gallica de la BNF et des écrits de Gilles Feyel.

comme un produit soumis à des impératifs d'utilité, de désirabilité et de viabilité (Tim Brown, 2014, 21).

Cela n'empêche pas Adolphe Thiers et Armand Carrel de fonder *Le National* en 1830, dans la pure tradition du journal d'opinion politique et sans publicité. Organe de promotion de la monarchie représentative, Thiers y prône la limitation du pouvoir du Roi le 4 février 1830. Plus tard, *Le National* d'août 1848 annonce un réel projet journalistique érigeant la presse comme « un gouvernement de pensée à côté du gouvernement de fait » (Feyel, 2003). En contrepied de la ligne partisane du *National* se trouve *La Presse* d'Émile de Girardin qui signe ce manifeste dans le numéro de lancement le 15 juin 1836 (voir Figure 1).

Figure 1. Extrait de « L'esprit de la rédaction » de Émile de Girardin à la une du numéro de lancement du 15 juin 1836

LA PRESSE différera des journaux établis principalement en ces points :  
 Que le prix de son abonnement ne sera point une spéculation ;  
 Qu'elle ne sera l'exploitation vénale d'aucun parti passionné ;  
 Qu'elle ne sera ni obséquieuse ni agressive envers le Pouvoir, mais juste et vraie ;  
 Qu'il ne sera point nécessaire d'être sacramentellement affilié à sa rédaction pour user de sa publicité en faveur d'une idée dignement exposée, d'un appel généreux, d'une relation exacte, d'une opinion portant la garantie d'une signature honorable.  
 On a défini la Presse : « La grande tribune des temps modernes. » Cette définition est fautive.  
 Les opinions contraires, diverses, généralement ou partiellement divergentes se succèdent à la tribune, elles s'excluent dans un journal.

Source : gallica.bnf.fr / Bibliothèque nationale de France

Antithèse du *National*, il prône une information apolitique et courtoise avec le « Pouvoir ». De Girardin entérine le modèle économique reposant sur les annonces pour créer un produit accessible et concurrentiel : « le journal est vendu une première fois aux annonceurs, avant de l'être aux abonnés », résume Feyel (2006).

Deux modèles économiques et visions du journalisme s'opposent. *Le National* promeut un modèle politique alternatif, embarquant ses lecteurs dans le débat quand *La Presse* prône une information rigoureusement neutre – sans se cacher d'assurer des débouchés commerciaux à l'industrie. Or l'impératif de satisfaire les annonceurs au détriment de l'émulation politique déstabilise les rédactions et la qualité du débat public. Tant et si bien qu'à la fin du XIX<sup>e</sup> siècle, Gustave Le Bon dans *Psychologie des foules* constate que « [l]es vieux organes solennels et

influents d'autrefois, dont la précédente génération écoutait pieusement les oracles, ont disparu ou sont devenus feuilles d'informations encadrées de chroniques amusantes, de cancans mondains et de réclames financières » (Le Bon, 1995, 88). Le Bon pointe également un renversement du rapport de force : la presse « autrefois directrice d'opinion » s'efface devant « l'opinion populaire » en donnant aux foules ce qu'elles demandent pour ne pas perdre leur lectorat.

## **2.2. Le temps des professionnels**

Conscient du pouvoir des foules jusqu'à parler de « tyrannie », l'américain Walter Lippmann propose de les remettre en place « pour que chacun puisse vivre sans craindre le rugissement et le piétinement du troupeau désorienté » (Lippmann, 1993, 145). Avec *Public Opinion* (1922) et *The Phantom Public* (1925), Lippmann souligne les limites d'une démocratie misant démesurément sur un « citoyen omniscient » (Lippmann, 2009, 273). Pour évacuer les stéréotypes du citoyen *lambda*, Lippmann assume que la production et la diffusion de l'information doivent être organisées par un corps de journalistes professionnels, « une classe spécialisée » (Lippmann, 2009, 310) dont le devoir est d'exposer le monde tel qu'il est. Pour Christopher Lasch, le projet de Lippmann a profondément déterminé le journalisme moderne : un « journalisme [...] guidé par le nouvel idéal de l'objectivité » (Lasch, 2007, 174).

Le projet de Lippmann ne s'arrête pas à la réforme du journalisme. Il souhaite aussi repenser la gouvernance de ce qu'il appelle la *Great Society* contemporaine en incitant les élites médiatiques et politiques à puiser dans la science et l'ingénierie. Entre autres professionnels de la mesure et de l'analyse quantitative, Lippmann s'attarde particulièrement sur le « scientifique social » capable de « rassembler ses données depuis une masse de matière hétérogène » (Lippmann, 2009, 374) pour analyser la complexité du monde et rationaliser la prise de décisions des « hommes d'action ». Cette captation des faits sociaux, leur analyse et leur restitution préfigurent l'emploi de méthodes statistiques pour étudier la société depuis des données brutes – nous y reviendrons.

Malgré la formalisation de la profession, le journalisme atteint la fin du XX<sup>e</sup> siècle en pleine « crise de confiance et de légitimité » (Schön, 1983, 11). Il est catégorisable dans ces professions de l'industrie de la connaissance jugées par le public comme ayant échoué à relever les enjeux contemporains. Il aurait failli à son devoir d'information, servi d'autres intérêts que l'intérêt public, entravé plus que soutenu les mécanismes démocratiques. Ce désajustement entre l'attente du public d'une information « zéro défaut » et les journalistes travaillant au mieux sans « prétendre à une telle objectivité » est la source d'un « grand malentendu » (Charon, 2007, 9). Pour éduquer le public aux médias tout en pensant de nouvelles entités journalistiques, il faudrait éclairer ce que Pierre Bourdieu appelait « l'emprise du journalisme » (1994), à savoir l'influence du marché ou encore des sondages sur le « champ journalistique » – que nous appellerons cadre de la production de l'information.

### **2.3. Distorsion ou construction ?**

Aussitôt conceptualisée, l'objectivité journalistique s'est entachée de nombreux biais imputés tant à la nature humaine du journaliste qu'à son environnement complexe. L'étude de Noam Chomsky et Edward Herman, publiée sous le titre de *La Fabrication du Consentement*<sup>2</sup> en 1988 éclaire de nombreux « filtres » modulant l'information entre sa collecte et sa publication : la structure financière du média, l'influence de la publicité, le poids des sources d'information primaires (sources « officielles », agences et experts), les divers moyens de pression sur le média, la construction d'un ennemi commun (communisme, terrorisme, ultra-gauche) (Chomsky et Herman, 2008, 26 sq.). Au lieu de penser l'information comme biaisée dans son objectivité (distorsion), l'analyste des médias Robert A. Hackett propose une approche holistique des conditions de production de l'information (construction). Hackett cite le sociologue Todd Gitlin qui définit ces « cadres de l'information » (*news frames* ou *media frames*) comme des « *patterns* persistants de cognition, d'interprétation et de présentation, ou encore de sélection, d'emphase et d'exclusion [qui] organisent au quotidien le discours, de manière verbale

---

<sup>2</sup> Terme emprunté à Walter Lippmann (voir Lippmann, 2009, 248).

ou non verbale » (Gitlin, 1980, 7). Le journaliste produit l'information selon une matrice de pensée propre à son journal et à ses représentations sociologiques. Son travail est structuré et orienté par tout ce qui « façonne le fonctionnement des médias comme institutions idéologiques » (Hackett, 1984, 254) :

*[L]e favoritisme partisan ou les préjugés politiques [...] les critères de sélection des informations, les caractéristiques technologiques des médias, les contraintes logistiques et budgétaires ou légales, la disponibilité des sources d'information, le besoin de raconter des histoires intelligibles et divertissantes à une audience ciblée, le besoin de packager l'information de façon compatible avec les impératifs commerciaux des annonceurs, l'apparition d'événements sociaux ou politiques. (Hackett, Ibid.)*

Cette vision constructiviste de l'information rapproche le travail de journaliste du travail de designer, au sens où l'information doit faire l'objet de recoupements et de composition avant d'être présentée à un destinataire. Comme l'énonçait la chercheuse en communication Brenda Dervin : « L'information n'a rien de naturelle. L'information – qu'on l'appelle donnée, connaissance, fait, chanson, histoire ou métaphore – est toujours designée [*designed*]. » (Dervin, 1999, 36) Or si le design est souvent invoqué dans la presse comme pratique graphique, il questionne plus largement les formats, les modes de lecture voire d'interaction, mais aussi le processus par lequel est produite l'information.

Cette vue globale du design sur la production de l'information lui a permis de rapidement s'adapter au *data deluge* (Anderson, 2008) des années 2000. Il s'est accoutumé à travailler avec les données et ses sciences et techniques afférentes, trouvant des applications à ces informations en devenir. Par convergence disciplinaire du design, de l'informatique et du journalisme traditionnel, le journaliste a désormais accès à ces techniques et manières de produire l'information.

### 3. La donnée : ressource et argument

#### 3.1. (Im)précision sur le journalisme de données

Biais de la « société de l'image » (Faccioli, 2007), le journalisme de données – comme le design de données – est souvent réduit à la traduction des données en images : on parle de visualisation de données ou encore de *graphic news* (qui les situent dans le champ des médias). Composé de deux termes eux-mêmes difficiles à définir, certains de ses adeptes ironisent à son propos : « Qu'est-ce que le journalisme de données ? Je pourrais répondre, simplement, que c'est faire du journalisme avec les données. » (Gray *et al.*, 2012, 2) D'autres terminologies précisent notre objet : le journalisme d'investigation numérique (*digital investigative journalism*) serait un journalisme tirant parti des techniques numériques, parfois même exclusivement « axé sur la donnée » (*data-driven journalism*) (Hahn et Stalph, 2018) ; le *nerd journalism*, observé par Alberto Cairo Tourinõ (2017) au sein de la rédaction de ProPublica, caractérise sociologiquement les personnes qui y travaillent et souligne leur appétence pour la création et l'expérimentation de nouveaux outils numériques pour traiter et représenter l'information.

Ce journalisme requiert une compréhension profonde des données, de leur « chaîne de valeur<sup>3</sup> » et des compétences permettant de les manipuler. Il regroupe de nombreux praticiens (journalistes, informaticiens, designers ou encore juristes) participant au processus par lequel les données brutes se transforment en information intelligible puis en connaissance appropriable par le public<sup>4</sup>. Sa finalité est de prolonger par de nouvelles techniques le dessein originel du journalisme : informer utilement le public.

Or entre la donnée et l'information utile et intelligible, il y a tout un ouvrage. Cela implique d'abord d'*agréger* les données adéquates : base de données, données ouvertes, etc. Il s'agit ensuite de les *nettoyer*

---

<sup>3</sup> Usuellement : captation, stockage, nettoyage, traitement, restitution.

<sup>4</sup> Le modèle « donnée → information → connaissance → sagesse (*wisdom*) » a été proposé par Nathan Shedroff (1999, 271).

(supprimer ou compléter les données partielles par exemple) et de les formater pour les rendre exploitables par l'humain ou l'algorithme. On peut ainsi les *analyser* avec un outil simple (tableur) ou complexe (algorithme statistique et apprentissage automatique). Enfin, ces résultats sont soumis à une médiation pour les *restituer* de façon intelligible au commun des mortels voire les rendre interactifs si l'information produite est manipulable selon plusieurs points d'entrée.

Comme de nombreuses disciplines des SIC et du design, le journalisme de données est plus facile à définir en termes de projets, d'outils et d'acteurs.

Un évènement comme Cambridge Analytica (2018) témoigne de son potentiel. En amont de l'affaire (2017), une équipe de chercheurs du Tow Center for Digital Journalism avait repéré dans le dépôt public Github<sup>5</sup> d'un développeur (*data scientist*) de Cambridge Analytica un fichier contenant des variables telles que « voterID », « tweetID » ou encore « sentiment ». L'investigation a montré que « les réactions émotionnelles des utilisateurs de Twitter et les discussions, favoris et likes sont extraits en temps réel et utilisés pour générer [...] des mots clés en rapport avec des campagnes politiques » (Albright, 2017).

Des outils comme le Global I-Hub (plateforme de rédaction virtuelle) ou encore l'Offshore Leaks Database (explorateur des Panama & Paradise Papers) assistent et facilitent le métier et la collaboration des journalistes.

Du côté des acteurs, certaines rédactions intègrent pleinement le paradigme numérique et le travail des données à leurs méthodes : le site de l'ONG américaine ProPublica s'est doté d'un onglet *Data & Graphics*<sup>6</sup> proposant de « Nouvelles applications, des graphiques, des bases de données, des outils » ; l'équipe pluridisciplinaire des Décodeurs

---

<sup>5</sup> Github est une plateforme d'hébergement et de versionnage de projets numériques.

<sup>6</sup> Voir en ligne : <https://www.propublica.org/newsapps/>

du *Monde* s'est dotée d'une charte<sup>7</sup> intégrant explicitement le journalisme de données.

### 3.2. De la donnée à l'information

À la lecture des pionniers de l'architecture et du design d'information tels que Richard Wurman ou Nathan Shedroff, l'environnement quotidien qui « bombarde nos sens » (Shedroff, 1999, 270) ne nous fournit pas de l'information mais de la donnée. En effet, « l'âge de l'information » (Mattelart, 2000) est concomitant d'une « explosion de la non-information » (Wurman, 2000). Pour prétendre au titre d'information, cette matière doit faire sens et pouvoir se communiquer. Il faut raffiner les données pour obtenir de l'information : « Pour avoir de la valeur, elles doivent être organisées, transformées et présentées de façon à leur donner du sens [...] » (Shedroff, 1999, *Ibid.*) La donnée ne se restreint pas à la cellule d'un tableur ; tout ce qui n'est pas intelligible, utilisable, exploitable, peut être taxé de *donnée*<sup>8</sup>. La frontière entre *donnée* et *information* devient ténue, si bien que les spécialistes du design d'information comme Edward Tufte<sup>9</sup> alternent *data*, *quantitative information* ou *complex information* (Tufte, 1983) pour désigner la matière brute « en entrée » du processus de design.

Plus complexe encore, l'information d'un média peut devenir la donnée d'un autre. Ce qui advient dans les visualisations de données « ALLÔ PLACE BEAUVAU ? C'est pour un bilan (provisoire) » nourries du recensement de violences policières mené par David Dufresne depuis

---

<sup>7</sup> Voir en ligne :

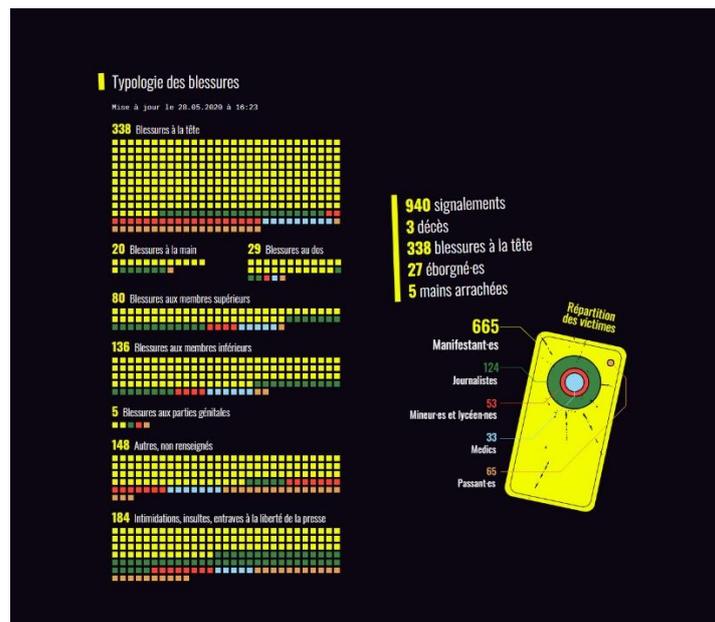
[https://www.lemonde.fr/les-decodeurs/article/2014/03/10/la-charte-des-decodeurs\\_4365106\\_4355770.html](https://www.lemonde.fr/les-decodeurs/article/2014/03/10/la-charte-des-decodeurs_4365106_4355770.html)

<sup>8</sup> On peut penser aux fuites (*leaks*) qui sont souvent des documents comptables ou courriers électroniques, peu compréhensibles sans explication et/ou mise en contexte. C'est pourquoi certains médias en vulgarisent le contenu et les enjeux, voir par exemple le Hors-Série du *Monde* n°24 intitulé « Le meilleur de WikiLeaks » dirigé par Sylvie Kauffmann en 2011.

<sup>9</sup> Edward Tufte est professeur de statistique et de design d'information à l'Université Yale.

novembre 2018<sup>10</sup> (voir Figure 2). Les tweets du compte de Dufresne « sont importés automatiquement dans un tableur confidentiel » (Dufresne, 2019) augmenté de métadonnées. Une fois complétés et vérifiés, les signalements intègrent la base de données qui actualise les visualisations : une histoire détaillée (information) devient un point sur un des graphiques (donnée) rapportant le phénomène des violences policières (connaissance). Dufresne et ses collaborateurs étant pionniers du format interactif, la page dédiée de *Mediapart* articule l'échelle qualitative et quantitative de l'information. Dans ce cas la réversibilité entre le format de donnée et d'information est un parti pris de design – un jeu favorisé par l'interopérabilité et l'interactivité des techniques numériques.

Figure 2. Typologie des blessures, issue de ALLÔ PLACE BEAUVAU ?  
C'est pour un bilan (provisoire).  
Avec l'aimable autorisation de David Dufresne.



Crédits : David Dufresne / Etamin Studio / Visionscarto / Mediapart

<sup>10</sup> Voir en ligne : <https://www.mediapart.fr/studio/panoramique/allo-place-beauvau-cest-pour-un-bilan>

### 3.3. Les biais des données

La transformation des données en information est un processus complexe qui recèle de nombreux pièges théoriques et pratiques. Edward Tufte cite les problèmes inhérents à la donnée ou à son approche – « une théorie stupide induit un graphique stupide » (1983, 15) – mais s'intéresse aussi et surtout aux biais de design. Au-delà des polarités data/design, on peut circonscrire des étapes sensibles de la chaîne de valeur de la donnée : la récupération, le traitement et la restitution.

Premièrement, la récupération des données conditionne la matière première de l'investigateur et influence jusqu'aux modes de restitution de l'information : « La nature des sources qui composent la visualisation est un critère qui détermine la nature du produit fini lui-même. » (Grandjean, 2015, 118) On évalue également si les données sont suffisantes et/ou comparables entre elles : un jeu de données peut s'avérer fragmentaire ou obsolète ; les données peuvent être difficiles à comparer, comme l'a montré la disparité de comptage des décès entre les différents pays durant l'épidémie de Covid-19 en 2020. On peut enfin questionner en amont la source même de ces données, susceptible d'édulcorer la matière fournie de ses « arrières-pensées politiques » (Laguës, 2014) – ce qui recoupe le troisième filtre de Chomsky et Herman sur les sources d'information primaires. Dans la même tonalité, le journaliste Paul Bradshaw avertit le néophyte quant aux données « prêtes à être analysées » qui, sans recul critique, conditionnent jusqu'au « type d'histoires que nous sommes à même de raconter » (Hahn et Stalph, 2018, 21).

Deuxièmement, le temps de l'analyse comporte aussi son lot de biais, soit dans l'approche théorique du jeu de données, soit dans l'emploi des techniques de traitement de données. L'abondance de données et la puissance computationnelle pour les traiter ne dispense pas d'une formalisation théorique solide. Edward Tufte se méfie des formulations théoriques flattant les discours et *a priori* des investigateurs : « Si vous torturez les données assez longtemps, elles vous révéleront n'importe quoi<sup>11</sup>. » (2016) Ne cherchant que ce qu'il veut voir, le commanditaire de

---

<sup>11</sup> Cette citation est attribuée au statisticien John W. Tukey.

l'algorithme risque d'établir des liaisons arbitraires entre certains points. Trouver des corrélations sans cause est d'ailleurs une caractéristique des algorithmes statistiques et d'apprentissage automatique. S'il fallait hier entrer des données dans un modèle théorisé pour obtenir un résultat, on délègue aujourd'hui la conception du modèle à l'algorithme auquel on fournit les données d'entrées et les données espérées en sortie – ce qui fit décréter à Chris Anderson la « fin de la théorie » en 2008 dans le magazine *Wired*. Le sociologue Dominique Cardon illustre avec facétie ce genre d'incohérence de la « logique » algorithmique :

*Il est par exemple possible que, selon les profils, la prédiction d'aimer la chose A dépende dans un cas plutôt de la couleur des yeux, de l'origine sociale et du nombre de déménagements et, dans un autre cas, du fait d'avoir voyagé en Estonie et d'avoir lu les œuvres complètes de Balzac. (Cardon, 2014)*

Le *data-driven journalism* devrait être interprété comme un journalisme aiguillé ou étayé par les données, non comme un journalisme aveuglément piloté par la technique. Loin d'obtenir de l'information depuis un processus automatique, l'humain sera toujours requis pour « regarder des données brutes souvent déroutantes ou ennuyeuses et “voir” les histoires cachées à l'intérieur » (Gray *et al.*, 2012, 5).

Troisièmement, nous arrivons justement à la restitution de ces données, et ce sous d'innombrables formats : l'article d'Albright sur Cambridge Analytica prend une forme rédactionnelle « classique » ponctuée de fragments de codes, liens et captures d'écrans ; le travail de Dufresne est distribué dans un dispositif médiatique couplant (entre autres) son compte Twitter et la page dédiée sur *Mediapart*. La restitution graphique n'est pas exclusive mais souvent employée. Les préceptes du design d'information sont alors requis pour organiser et hiérarchiser les données jusqu'à restituer une information intelligible. Les journalistes et designers doivent faire preuve à ce stade d'intégrité graphique comme dirait Tufte (1983, 53 *sq.*). Le design d'information doit focaliser l'attention sur la substance du graphique, permettre de remarquer une singularité dans un ensemble, ne pas dévoyer les données par des variations de design. Mais Tufte relève aussi dans *The Visual Display of Quantitative Information* de nombreux moyens de faire mentir les données par le design : un dessin en perspective simulant une croissance

ou un jeu d'échelle majorant un phénomène. Heureusement, notre « détecteur de mensonge graphique » (Tufté, 1983, 53) sait repérer les données malmenées. Début avril 2020, un graphique de la chaîne Fox 31 devenait viral sur Internet : les nouveaux cas de Covid-19<sup>12</sup> y étaient représentés en ordonnées sur une graduation dont l'intervalle oscillait follement entre 30, 10 ou 50 (voir Figure 3). Bien que le pic entre les 25 et 26 mars soit quelque peu atténué par cette variation, les internautes penchent pour de l'incompétence plutôt que de la malveillance.

Figure 3. Diagramme illustrant le nombre quotidien de nouveaux cas de Covid19 aux États-Unis entre le 18 mars et le 01 avril 2020 sur la chaîne Fox 31.



Sources diverses : Reddit, Twitter.

Nous avons évoqué le fait que la visualisation graphique est un mode récurrent de restitution, sans pour autant être exclusif. Il faut alors poser la question du bien-fondé de son usage au regard de la situation (format des données, phénomène observé, destinataire). L'information doit apparaître au grand jour sans être « noyée dans les nombreuses statistiques présentées » (Laguës, 2014). Jacques Bertin dans sa *Sémiologie graphique* questionne son lecteur en ces termes : « Faut-il faire un dessin ? Suivant le cas, le dessin peut être inutile ou nécessaire. »

<sup>12</sup> Voir en ligne (sources diverses) :

<https://twitter.com/Carnage4Life/status/1246579721585868800> ou :

[https://www.reddit.com/r/CrappyDesign/comments/fv0fpz/the\\_yaxis\\_on\\_this\\_fox\\_news\\_graph\\_of\\_coronavirus/](https://www.reddit.com/r/CrappyDesign/comments/fv0fpz/the_yaxis_on_this_fox_news_graph_of_coronavirus/)

(1999, 100) La visualisation de données doit donc être employée pour un gain d'intelligibilité, non pour son esthétique légitimante.

### **3.4. Vers une rhétorique des données ?**

Bien qu'elles étayent indubitablement les ressources du journaliste, les données utilisées à l'excès font courir à la profession le risque d'une fatuité technique plus dommageable que bénéfique. Ce risque est la systématisation de la production d'information depuis la donnée, instaurant dans son paroxysme un mode de production qui ne serait plus initié ni médié par le journaliste humain. Des expériences d'algorithmes d'écriture automatique agencant des données en formats narratifs ont ouvert la voie à l'automatisation de la production d'information : le Quakebot du *Los Angeles Times* (2014) rédige des alertes aux tremblements de terre, l'Heliograph du *Washington Post* (2017) des actualités sportives, le Cyborg de Bloomberg (2018) des actualités financières.

Sans en arriver à cette extrême robotique, le journalisme peut être tenté de se légitimer par les données, usant de leur poids rhétorique à toute échelle : une matière pure (source) est travaillée par un processus implacable (traitement algorithmique) produisant une visualisation péremptoire (restitution). L'éthique du journaliste de données consiste au contraire à éviter l'analogie trop confiante entre ce qui se trame dans l'ordinateur et ce qui advient dans nos vies. Rappelons que les données demeurent des instantanés de comportement ou de phénomène arrachés à leur contexte par la conversion de l'analogique au numérique. La juriste et technologue Antoinette Rouvroy relie justement leur caractère implacable à leur production automatique :

*Une donnée n'est plus qu'un signal expurgé de toute signification propre – [...] ce qui semble assurer leur prétention à la plus parfaite objectivité : aussi hétérogènes, aussi peu intentionnées, tellement matérielles et si peu subjectives, de telles données ne peuvent mentir !*  
(Rouvroy et Berns, 2013)

On observe ainsi au cours de l'Histoire une superposition des garanties de l'objectivité. *La Gazette* de Renaudot y prétend de par son réseau de correspondants eurasiens et son adoubement par le pouvoir

politique. De Girardin explique n'ouvrir les colonnes de *La Presse* qu'aux « signatures honorables », préfigurant les « professionnels de l'information » que décrit Lippmann. Ce dernier jette en 1922 un pont entre l'ancien monde et le nouveau, en plaçant les sciences et techniques en support des élites médiatiques et politiques. L'enregistrement massif des faits sociaux donnerait aux « scientifiques sociaux » la matière pour analyser, dire et prédire le monde. Mais cette pratique que Rouvroy appelle la « mise en nombres de la vie » (Rouvroy et Stiegler, 2015) ne garantit en rien la pertinence de la réinjection du produit du calcul dans le monde. Le recours aux bases de données, aux algorithmes, puis à la visualisation de données devrait se pratiquer en conscience et connaissance des multiples biais évoqués. Et si la donnée se présente comme une modalité supplémentaire de l'argumentation journalistique, elle ne doit pas en devenir l'argument exclusif et suffisant, sous peine de lire nos existences au seul prisme de « l'uniformité statistique » (Arendt, 2002, 82).

*« Comme toute source, [la donnée] doit être traitée avec scepticisme ; et comme tout outil, nous devons être conscients de la façon dont elle peut façonner et restreindre les histoires qu'elle permet de raconter. » (Gray et al., 2012, 3) Ce qui se joue dans ce rapport à la donnée, c'est donc la capacité des journalistes à construire des histoires signifiantes pour le public. Notre lecture des données s'apparente au tracé des constellations d'étoiles en étoiles, « [c]'est-à-dire que les points en eux-mêmes ne veulent pas dire grand-chose mais que les GENS et les ordinateurs les relient pour écrire leurs propres histoires » (Keller et Neufeld, 2017, 19).*

#### **4. Conclusion**

Cette réflexion a montré que les données et leurs techniques afférentes modifient les manières de faire des journalistes et influencent le cadre de la production de l'information. Pour autant, l'introduction de la donnée n'ébranle en rien les fondements du journalisme ; elle s'inscrit dans la continuité de pratiques existantes et bien ancrées. Le journalisme de données alimente en réalité le journalisme dans son ensemble :

*Une distinction binaire entre le journalisme numérique et analogique est en train de disparaître. Les nouvelles technologies de*

*communication sont désormais un élément essentiel du journalisme et prolongent des formats classiques du répertoire journalistique [...] (Hahn et Stalph, 2018, 2)*

Néanmoins, cette ressource pour l'activité journalistique implique une réflexion éthique sur les potentiels et contraintes des données. Les techniques numériques associées aux données viennent soutenir l'investigation (méthodologie) et pas seulement légitimer un discours par une autorité technique (rhétorique). Le journaliste demeure responsable, médiateur et vulgarisateur des données qu'il recueille, des algorithmes qui les traitent, des images qui les donnent à voir.

La culture technique qui pénètre le journalisme pourrait aussi amener un changement méthodologique. Le journalisme de données utilisant lui-même l'arsenal technique qu'il analyse, il peut inciter ses praticiens à vulgariser et expliciter leurs pratiques. Comme un développeur documente son code pour sa communauté, le journaliste documenterait sa méthodologie d'enquête pour une plus grande transparence au regard du public et de ses pairs. Certains journalistes ont déjà l'habitude d'exposer leur méthodologie d'investigation : Jonathan Albright documente son « analyse » du code de Cambridge Analytica tandis que David Dufresne explique le processus allant du premier repérage d'une violence policière à l'actualisation effective des graphiques sur *Médiapart*. Au vu de la matière brute, opaque et complexe manipulée, cette transparence n'est pas superflue.

Enfin, comme il comprend et s'approprie les paradigmes numériques, le journalisme de données est une des seules disciplines à même de décrypter l'écosystème complexe de l'industrie numérique s'immiscant toujours plus loin dans la sphère politique (affaire Cambridge Analytica, position monopolistique des géants du numérique). Inversement, il peut observer les mécaniques d'intrusion des États dans la vie des particuliers (révélations du projet PRISM, dysfonctionnement du STIC français). Fort de sa culture technique, capable d'explorer les obscures bases de données et lignes de codes, le journalisme de données peut agir en garde-fou des déviances de la société numérique en rendant intelligible ses rouages et implications sur nos vies.

## Bibliographie

- Albright Jonathan (2017). *Cambridge Analytica: The Geotargeting and Emotional Data Mining Scripts*,  
<https://medium.com/tow-center/cambridge-analytica-the-geotargeting-and-emotional-data-mining-scripts-bcc3c428d77f>
- Anderson Chris (2008). *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, <https://www.wired.com/2008/06/pb-theory/>
- Arendt Hannah (2002). *Condition de l'homme moderne*, Pocket, Paris.
- Bertin Jacques (1999). *Sémiologie graphique : Les diagrammes, les réseaux, les cartes*, EHESS, Paris.
- Bourdieu Pierre (1994). L'emprise du journalisme. *Actes de la recherche en sciences sociales*, vol. 101-102, pp. 3-9.
- Brown Tim (2014). *L'esprit design*, Pearson Education, Londres.
- Cardon Dominique (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Seuil, Paris.
- Cairo Touriño Alberto (2017). *Nerd journalism: How data and digital technology transformed news graphics*. Thèse en Société de l'Information et de la Connaissance. Universitat Oberta de Catalunya.
- Charon Jean-Marie (2007). *Les journalistes et leur public : le grand malentendu*, Vuibert, Paris.
- Chomsky Noam, Herman Edward (2008). *La Fabrication du consentement. De la propagande médiatique en démocratie*, Agone, Marseille.
- Dervin Brenda (1999). Chaos, Order and Sense-Making. *Information Design*, Cambridge, Londres, The MIT Press, pp. 35-57.
- Dufresne David (2019). *Six mois d'« Allô Place Beauvau » : chronique des violences d'État*,  
<https://www.mediapart.fr/journal/france/170519/six-mois-d-allo-place-beauvau-chronique-des-violences-d-etat>
- Faccioli Patrizia (2007). La sociologie dans la société de l'image. *Sociétés*, vol. 1, n° 95, pp. 9-18.
- Feyel Gilles (2003). Presse et publicité en France (XVIIIe et XIXe siècles). *Revue historique*, vol. 4, n° 628, pp. 837-868.
- Feyel Gilles (2006). Le coût de la presse quotidienne parisienne (1800-1844), *Le Temps des médias*, vol. 1, n° 6, pp. 9-21.
- Gitlin Todd (1980). *The whole world is watching*, University of California Press, Berkeley.

- Keller Michael et Neufeld Josh (2017). *Dans l'ombre de la peur : Le Big Data et nous*, Ça et là, Bussy-Saint-Georges.
- Hackett Robert A. (1984). Decline of a paradigm? Bias and objectivity in News Media Studies. *Critical Studies in mass Communication*, vol. 1, n° 3, pp. 229-259.
- Hahn Oliver, Stalph Florian (2018). *Digital Investigative Journalism, Data, Visual Analytics and Innovative Methodologies in International Reporting*, Palgrave Macmillan, Londres.
- Grandjean Martin (2015). Introduction à la visualisation de données : l'analyse de réseau en histoire. *Geschichte und Informatik*, vol. 18 et 19, pp. 109-128.
- Gray Jonathan, Bounegru Liliana, Chambers Lucy (2012). *The Data Journalism Handbook*, O'Reilly, Sebastopol.
- Laguës Benjamin (2014). *Le data journalism, miroir aux alouettes ou renouvellement du métier ?*, <https://www.acrimed.org/Le-data-journalism-miroir-aux-alouettes-ou-renouvellement-du-metier>
- Lasch Christopher (2007), *La révolte des élites et la trahison de la démocratie*, Flammarion, Paris.
- Le Bon Gustave (1995). *Psychologie des foules*, PUF, Paris.
- Lippmann Walter (2009). *Public opinion*, Transaction Publishers, New Brunswick, Londres.
- Lippmann Walter (1993). *The Public Phantom*, Transaction Publishers, New Brunswick, Londres.
- Mattelart Armand (2000). L'âge de l'information : genèse d'une appellation non contrôlée. *Réseaux*, vol. 18, n° 101, pp. 19-52.
- Renaudot Théophraste (1631). Préface. *La Gazette*, Bureau d'adresse, Paris.
- Rouvroy Antoinette, Berns Thomas (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? *Réseaux*, vol. 1, n° 177, pp. 163-196.
- Rouvroy Antoinette, Stiegler Bernard (2015). Le régime de vérité numérique. *Socio*, vol. 4, pp. 113-140.
- Schön Donald (1983). *The Reflective Practitioner*, Basic Books, New York.
- Shedroff Nathan (1999). Information Interaction Design: A Unified Theory of Design. *Information Design*. Cambridge, Londres, The MIT Press, pp. 267-292.
- Tufte Edward (1983). *The Visual Display of Quantitative Information*, Graphic Press, Cheshire.
- Tufte Edward (2016). *The Future of Data Analysis* (conférence). Microsoft Machine Learning & Data Science Summit 2016.

Wurman Richard Saul (2000), *Information Anxiety 2*, QUE, Indianapolis.