

Olivier Bot

J⁺ournalisme

Chercher
et enquêter
avec Internet

Préface de Johan Hufnagel

PUG

Olivier Bot

Chercher et enquêter avec Internet

Presses universitaires de Grenoble

Introduction

Depuis sa démocratisation à la fin des années 1990, Internet a changé le rapport entre société et information. La progression de la couverture planétaire du wifi et celle de la télévision par satellite ont connecté des sphères d'actualité jusqu'ici séparées. En 2005, une série de caricatures de Mahomet publiées dans le quotidien danois *Jyllands-Posten* a embrasé le monde musulman. Désormais tout fait écho sur Internet et tout fait lien, provoquant un choc des cultures.

Le Web a aussi permis à tout un chacun de transmettre un fait ou une opinion par un canal direct, reléguant ainsi la médiation journalistique au rang d'une expression parmi d'autres, si ce n'est qu'elle garantit toujours des nouvelles certifiées. Internet a donc profondément transformé l'information et sa portée, assignant au journaliste un rôle social plus modeste et une responsabilité plus grande. Dans ce même mouvement de transformation, Internet a ouvert au journalisme de nouveaux espaces, lui permettant de mieux rendre compte de la complexité des faits et de la diversité des opinions.

Cette technologie nouvelle a aussi eu un impact sur les pratiques professionnelles, comme d'autres avant elle. Le télégraphe a produit au xx^e siècle le *lead* d'agence, l'information brute, celle qui doit passer coûte que coûte avant une possible coupure de ligne. Au xxi^e siècle, Internet a suscité le tweet de 140 signes, le microblogging, le journalisme de données, ou encore le long format, qui allie vidéo, texte, son, photographie et infographie, jusqu'ici portés par des médias différents.

« L'ADN de l'info a changé, il faut changer l'ADN des journalistes », écrivait Benoît Raphaël sur son blog « La social Newsroom » (benoitraphael.com). Si Internet a modifié la pratique du journalisme,

les rédactions ne s'y sont pas toutes adaptées et n'ont pas toutes développé une interactivité réelle avec leurs lecteurs, pourtant devenus des prescripteurs d'informations. Les journalistes n'ont pas encore tous acquis une connaissance profonde de la navigation sur le Web, ni utilisé les nouveaux outils digitaux qui sont à leur disposition. Nombre de sites d'information sont même tombés dans le *churnalism* (en reprenant par copier-coller des dépêches, des communiqués de presse ou des informations de seconde main, voire en plagiant des articles produits par d'autres (Davies, 2008)). Une étude menée sur un très important corpus d'articles de presse publiés en 2013, commenté dans le livre *L'information à tout prix* de Julia Cagé, Marie-Luce Viaud et Nicolas Hervé, a montré récemment que « 64 % des contenus journalistiques sont de purs copier-coller d'autres articles » (InaGlobal, 2017).

Le Web dans toutes ses dimensions offre pourtant de nouvelles opportunités : une puissance documentaire inégalée, un carnet d'adresses infini, une mémoire puissante, des moyens d'alerte rapide, une automatisation des tâches, des matériaux et des outils d'enquête pertinents ainsi que des réseaux d'interactivité étendus, tant dans l'élaboration des sujets que dans l'évaluation des articles. Mais la Toile tend aussi de nombreux pièges aux moins précautionneux. Les écueils sont les mêmes qu'hier, sauf que l'immédiateté et ce nouveau champ d'expérience encore mal maîtrisé ont multiplié les chausse-trappes.

La campagne électorale étasunienne de 2016 a aussi marqué un tournant dans la perception de l'information, de son rôle et de sa diffusion sur Internet. Pour la première fois, en effet, selon une étude de BuzzFeed¹, les internautes américains ont partagé et commenté plus d'informations fausses ou invérifiables que de nouvelles exactes et vérifiées, et le mot *post truth* (post-vérité) a fait une entrée fracassante dans le prestigieux dictionnaire britannique Oxford comme étant le mot de l'année 2016. Ces *alternative facts* (faits alternatifs), concoctés par l'administration Trump, constituent un défi au journalisme, opposant un cocktail d'émotions et d'opinions aux vérités factuelles.

1. « Ils ont hacké la politique » : <https://www.slideshare.net/fabienFICHET/hackerlapolitique>

De plus, les deux candidats à la présidence, Donald Trump et Hillary Clinton, ont tous deux utilisé des potentialités différentes de l'Internet pour mener leur campagne. Le républicain a fait campagne sur Twitter et la démocrate a utilisé des fichiers de données pour toucher les électeurs (Bussard, Mottaz, 2016). Pour rendre compte de l'élection, il fallait donc s'intéresser à leur stratégie sur Internet, et ne pas se laisser prendre aux *fake news* (fausses nouvelles).

La prolifération de ces fausses informations et la mise en ligne de données manipulées outre-Atlantique ont accéléré une prise de conscience, en France comme ailleurs, de la nécessité de trier systématiquement le bon grain de l'ivraie, tandis que les premières tentatives de certification de l'information étaient entreprises par différents acteurs du numérique. Ce qui n'est, soit dit en passant, pas sans danger pour la liberté d'expression.

Le réseau social Facebook, important vecteur de *fake news*, a pour sa part réagi aux critiques grandissantes de l'opinion en lançant son Journalism Project mettant des outils de vérification à disposition des journalistes. Avec d'autres mécènes, de grandes universités et d'autres acteurs de l'économie numérique, l'entreprise de Mark Zuckerberg a aussi cofinancé un fonds de quatorze millions de dollars pour lutter contre la désinformation². Sous pression, Facebook a même fini par modifier ses algorithmes afin de privilégier le cercle amical et l'information de proximité pour chaque membre du réseau.

Avec une Toile ouverte à tous les vents, l'information est devenue un combat de tous les instants. Car dans cet espace d'échanges, se côtoient décideurs, publicitaires, influenceurs, communicants, propagandistes, leaders d'opinion, amuseurs, hackers, cyberactivistes, journalistes et simples citoyens. Démêler le vrai du faux, chercher avec acuité, vérifier des affirmations et de supposés faits sont autant de tâches que les journalistes exerçaient hier dans le monde réel et dont les règles, si elles n'ont pas changé, doivent s'adapter à ce nouvel environnement.

2. Cambell B., 2 avril 2017, "Introducing the News Integrity Initiative": <https://media.fb.com/2017/04/02/introducing-the-news-integrity-initiative/> et Cunny J., <https://www.journalism.cuny.edu/2017/04/announcing-the-new-integrity-initiative/>

Le journaliste sait-il décrypter une URL ? Cibler sa demande ? Retrouver une archive du Net ? Vérifier l'authenticité d'une photo ? Exploiter des données pour en tirer une information ? Retrouver des pages d'un site fermé ? C'est à toutes ces questions, parmi d'autres, que va répondre ce manuel, qui compile un certain nombre de sites et d'outils digitaux utiles aux journalistes. Tous ont été testés au cours de l'été et de l'automne 2017. Néanmoins, l'univers numérique étant en constante mutation, il est possible que certains disparaissent ou deviennent payants, changent de nom ou d'architecture. C'est pourquoi, dans la mesure du possible, deux voire trois outils – presque tous gratuits – seront proposés pour chaque tâche. De plus, afin de s'entraîner à maîtriser ces outils, des exercices pratiques sont proposés tout au long de l'ouvrage.

Dans le premier chapitre, il sera question de l'architecture du Web, du nécessaire décodage des URL, des moteurs de recherche, de leurs algorithmes, des moteurs dédiés et des métamoteurs. On y trouvera aussi des outils qui permettent de trouver plus vite ce que l'on cherche.

Le deuxième chapitre sera consacré aux recherches documentaires et aux bases de données, le matériau nécessaire au journalisme 2.0, au travail d'enquête et au data journalisme.

Le troisième chapitre s'intéressera aux pièges d'Internet et aux moyens de les éviter. À l'heure de la post-vérité décrétée par le président des États-Unis, des trolls de la propagande russe, des théories conspirationnistes et des légendes urbaines revisitées, le journaliste se doit de savoir reconnaître de fausses nouvelles et d'avoir des outils qui lui permettent de vérifier l'authenticité d'une image, d'un tweet ou d'une affirmation.

Le quatrième et dernier chapitre explorera les questions d'éthique professionnelle dans ce nouvel environnement digital. Comment garantir la protection des sources ou le respect de la vie privée, identifier ses interlocuteurs et les plagiat sur le Web ? Autant de règles éthiques qui doivent trouver leur chemin sur la Toile.

En conclusion, nous suivrons heure par heure une journée de travail rédactionnel avec Internet.

Les pièges du Net

Le candidat Donald Trump est soutenu par le pape. Faux. Emmanuel Macron possède un compte au Bahamas. Faux. Des décalcomanies au LSD sont distribuées à la sortie des écoles. Faux. Le vrai nombre de victimes suite au passage de l'ouragan Irma sur les îles de Saint-Martin et Saint-Barthélémy nous est caché. Faux. Les traces des avions (*chemtrails*) dans le ciel sont des épandages de toxiques pour nous empoisonner. Faux. Ebola est une arme biologique créée par le Pentagone. Faux.

À l'heure d'Internet, les rumeurs, les théories conspirationnistes, les *fake news* (fausses nouvelles), les éléments de propagande idéologique (les post-vérités par exemple) pullulent sur nos écrans. L'accélération médiatique, suite à la création des chaînes d'information continue à la radio et à la télévision et à la démocratisation d'Internet, augmente le risque que des journalistes de bonne foi, mais peu rigoureux ou pas assez vigilants, ne reprennent et diffusent ces pseudo-nouvelles. Pour faire des clics, que ce soit à des fins financières, de canular ou d'influence, les dérives sensationnalistes se sont multipliées, devenant même la politique rédactionnelle de certains médias parodiques.

Une industrie de la fausse nouvelle s'est même créée, tandis que les sites conspirationnistes se sont multipliés depuis les années 2000. À Vélès en Macédoine, les frères Velkovski (les « Healthy brothers ») ont eu l'idée de créer des sites internet attractifs pour gagner de l'argent grâce à Google AdSense, le service de publicité de Google (qui offre aux États-Unis entre 0,22 et 0,50 dollar par clic). Eux et d'autres

jeunes macédoniens ont ensuite flairé le filon Trump. Dans l'effervescence de la précampagne puis de la campagne et depuis, ils ont multiplié les titres les plus racoleurs, avec souvent des connotations idéologiques pro-Trump, prorusses et antidémocratiques. « Le pape soutient Trump » ou « Les femmes de ménage de la Maison-Blanche applaudissent le départ d'Obama » ont été publiés par eux, sur des sites comme Worldpoliticus.com ou DonaldTrumpNews.co (Barbier, Minisini, 2017).

Les répercussions de ces fausses nouvelles ont été importantes. Certains observateurs affirment qu'elles ont influé sur le résultat de l'élection présidentielle américaine. Si cela n'est pas certain, il est en revanche sûr que l'information en a pâti. Pour preuve, la fausse information faisant d'Hillary Clinton une marchande d'armes pour le groupe État islamique (Daech en arabe) a été plus partagée que la vraie déclaration de Barbara Bush, l'ancienne première dame sous le mandat du président républicain George W. Bush, se demandant « comment les femmes pouvaient bien voter pour Trump ».

La rumeur, les théories complotistes, les fausses nouvelles diffusées pour peser sur une élection, ce n'est pas nouveau. Mais la puissance des réseaux sociaux et leur fonctionnement en partage ont fait grandir le phénomène, au point de saturer le Web.

Mais heureusement, Internet a aussi fourni les antidotes aux poisons et conduit à un surcroît de vigilance dans les rédactions, tandis que la vérification des faits, l'alpha de la profession, donne lieu dans de plus en plus de médias à une rubrique dédiée au *fact-checking*.

Pour s'assurer de la qualité d'une information lue sur le Web, il faut faire preuve de bon sens. D'abord, il faut identifier la source. Est-elle sûre ? Ensuite l'URL peut vous donner quelques renseignements sur sa crédibilité. S'il s'agit d'un article de journaliste, vérifiez si ce confrère a eu lui-même accès à des sources ou s'il ne fait que reprendre une information lue ailleurs.

Si cela provient d'une source inconnue... méfiance. Sur beaucoup de sujets, on peut recouper avec ses propres sources afin de vérifier la crédibilité de la nouvelle. On peut aussi observer quel retentissement

a cette information sur les médias en ligne. La fiabilité de la source, la crédibilité et l'authenticité de l'information, sa vérification par d'autres canaux que le Web (qui fonctionne beaucoup en boucle, par partage de contenus) restent les meilleures garanties de la qualité d'une information lue sur Internet. Si elle est porteuse d'émotion et, *a fortiori*, si elle est anxiogène, il faut redoubler de prudence. Quand c'est trop « beau » en termes d'infos, c'est que c'est peut-être faux.

Savoir reconnaître une rumeur

La rumeur est probablement la plus ancienne des *fake news*, et tout journaliste avant même l'apparition d'Internet y a été confronté au cours de son activité. Depuis les années 1960 et le livre d'Edgar Morin sur la rumeur d'Orléans (une pseudo traite des blanches dans un grand magasin possédé par un juif de la ville), ces légendes urbaines ont été étudiées par des sociologues, permettant d'en établir une sorte de typologie.

Jean-Bruno Renard, professeur de sociologie à l'université Paul-Valéry Montpellier 3 et Véronique Champion-Vincent, chercheuse au CNRS, se sont tous deux intéressés au phénomène des rumeurs pour le définir et nous aider à les débusquer. Dans plusieurs livres, ils ont établi une classification qui permet au journaliste de repérer une possible rumeur, ou en tout cas de se méfier. Présentées comme vraies, les rumeurs expriment nos peurs, nos tabous et nos espoirs. Une de leurs caractéristiques est de renaître sous de nouvelles variantes et de ne pas avoir de source identifiée. Elles nous effraient, nous intriguent, nous étonnent, nous mettent en colère.

Dans leur dernier livre (2014), les auteurs prennent l'exemple de la rumeur des 40 000 femmes venues des pays de l'Est pour « offrir des services sexuels aux spectateurs de la Coupe du monde de football en Allemagne ». Une pétition d'une ONG de lutte contre le trafic humain avait crédibilisé cette affirmation. Or, cette histoire circulait déjà lors de l'édition 2010 du même événement en Afrique du Sud. La répétition d'un fait, déplacé dans le temps et dans l'espace, est un indice de possible rumeur.

Dans leurs livres précédents (2002 ; 1992), les auteurs avaient démonté le processus de fabrication d'une rumeur : l'amplification, le déplacement, la reconstruction. C'est pourquoi bon nombre de rumeurs sont des histoires de comble. Elles sont en général alléchantes pour les journalistes, très excités par la nouvelle qui sort de l'ordinaire. « C'est trop beau », « incroyable », « tu as vu cette histoire ? » : autant de phrases entendues dans les rédactions, qui doivent susciter le scepticisme. *A fortiori*, s'il s'agit d'une histoire virale recueillie sur Internet.

Les auteurs ont aussi établi une typologie des rumeurs qui doit éveiller le doute chez le journaliste. Il y a les rumeurs alimentaires : pizza à l'urine, hamburgers à l'araignée, souris trouvée dans du Coca-Cola, etc. ; les rumeurs technologiques : la persistance rétinienne des morts assassinés (comme dans *Quatre mouches de velours gris*, le film de Dario Argento), le nageur avalé par un Canadair, le strip-tease de sa copine envoyé automatiquement sur Internet (comme dans *American Pie*), etc. ; les peurs criminelles : le LSD en tatouage pour les enfants, un hangar abandonné qui cache un site nucléaire ; le retour du sauvage ou de la barbarie : de grands félins aperçus, voire photographiés dans un environnement qui n'est pas le leur, le crocodile surgi des toilettes, un camp d'extermination créé par Obama, etc. ; et le surnaturel : les dames blanches, Nostradamus qui aurait prévu le 11 septembre, etc.

Bien souvent, une rumeur ne peut se contredire qu'en posant des questions à un spécialiste. Par exemple, un pilote de Canadair vous dira que la buse qui pompe l'eau dans un lac ne peut avaler un plongeur car son diamètre est trop modeste. De même, à propos de décalcomanies soi-disant distribuées aux écoliers, un toxicologue vous dira que le LSD ne se diffuse pas par les pores de la peau. On constate ainsi la nécessité de joindre un spécialiste lorsqu'une histoire entrant dans cette typologie apparaît sur la Toile.

Mais comme les rumeurs se transforment et naissent à partir d'une bribe de faits d'actualité, il faut sans cesse remettre à jour sa base de données. C'est ce que fait le site HoaxBuster (hoaxbuster.com) dont le nom est constitué du mot *hoax*, qui signifie canular (les anglophones parlent de *bomb hoax*, pour une fausse alerte à la bombe) et de *buster*, quelqu'un qui brise, qui casse quelque chose.

Le site Hoaxkiller (hoaxkiller.fr) propose, lui, un moteur de recherche par mots-clés et en texte intégral. On peut se tenir au courant des dernières rumeurs, chaînes et pourriels repérés sur Internet en laissant son adresse e-mail. Les alertes au virus informatique, les disparitions d'enfant, les personnes qui promettent le bonheur ou la guérison et les pétitions douteuses sont aussi dans son champ d'expertise. Enfin, ce site français, qui réalise une fiche très complète sur chaque cas, publie un classement des dix *hoax* du moment.

D'autres formes de *hoax* et de canulars, plus ou moins indolores, parcourent le Web. Le précurseur de la mystification digitale est l'américain The Onion (theonion.com/). Parmi les sites francophones qui s'en sont fait une spécialité, le plus connu est le site français Legoraf.fr (anagramme de *Le Figaro*). Il a un équivalent en Belgique, le site NordPresse.be (nordpresse.be/). Il existe aussi un site parodique africain, State Africa (stateafrique.com/).

Des journaux sérieux s'y sont laissés prendre. C'est le cas du quotidien algérien *Al-Hayat* qui reprenait un article parodique du Gorafi qui s'inquiétait de voir Marine Le Pen vouloir construire un mur sur la Méditerranée, comme celui que Trump veut prolonger à la frontière entre les États-Unis et le Mexique. Des politiques comme Christine Boutin ou Jean-Marie Le Pen ont aussi cité l'article du Gorafi en le prenant pour argent comptant.

Ces parodies impactent parfois l'actualité elle-même, quand par exemple, une ouvrière de Whirpool déclare : « Monsieur Macron. On est des ouvriers. On a les mains propres. Vous nous serrez la main ? ». Un article du Gorafi³¹ avait fait dire au candidat que « c'était plus fort que lui, quand il serrait la main d'un pauvre, il fallait qu'il lave ses mains ». Un montage vidéo venait étayer le propos, puisqu'on y voit le candidat s'essuyer les mains avant de monter dans sa voiture. L'image avait été détournée, et provenait d'une visite à des pêcheurs d'anguilles dans un étang du Midi³².

31. Article du Legoraf : <http://www.legoraf.fr/2016/06/01/emmanuel-macron-quand-je-serre-la-main-dun-pauvre-je-me-sens-sale-pour-toute-la-journee/>

32. Une quinzaine de ces sites parodiques ont été listés par *L'Express*. Voir note 52 pour le lien.

Table des matières

Préface	5
Introduction	11
Chapitre 1. Surfer en journaliste	15
Le Web n'est pas un grand sac	15
Le Web réinvente le journalisme	17
Une URL, ce n'est pas de l'hébreu	20
Délier la chaîne de caractères	20
Les TLD pour s'y retrouver	21
Les moteurs de recherche	24
Les algorithmes nous dirigent	25
Google domine mais il n'est pas seul au monde	28
Des moteurs qui respectent la confidentialité	31
Pourquoi se contenter d'un seul moteur?	34
Aller dans le Web profond	36
La recherche experte	37
Merci Monsieur Boole	38
Les outils de recherche	40

Chapitre 2. Le Web documentaire	43
Revue de presse électronique	44
Les agrégateurs de flux RSS, tout sur un écran	45
Des journaux et des news	46
Des sources digitales	50
Wikipédia, la bible?	50
Des mots et des langues	53
Des sources savantes à son bureau	55
Je sais qui vous êtes	65
Les <i>leaks</i> et les <i>papers</i>	69
Faire parler les chiffres et les images	70
Des données à la base	70
Le journalisme de données	74
Des images à légender	76
Utiliser les réseaux et les indicateurs d'opinion	82
Des gazouillis, des hashtags et des pages Facebook	83
Les réseaux sociaux ne sont pas des agences de presse	86
Forums et blogs	88
Chapitre 3. Les pièges du Net	91
Savoir reconnaître une rumeur	93
Les théories complotistes	96
Les <i>fake news</i> et la crise médiatique de la post-vérité	99
Les algorithmes de circulation de l'information	100
Le <i>fact-checking</i>	101
Chapitre 4. Une éthique pour le journaliste internaute	105
Anonymat et confidentialité	106
Identifier qui est derrière un site	107
Respecter la vie privée	109
Protéger ses sources	110

Chercher la vérité	114
Détecter le plagiat	114
Injure, diffamation et incitation à la haine	116
S’informer et se former aux nouveaux outils	117
 Conclusion. Internet au quotidien	 119
 Mémento des outils du Web	 125
 Corrections des exercices	 131
 Bibliographie	 133