

# **Prises de positions et influences sur le web : le cas de l'information de santé.**

Amar LAKEL  
Laboratoire MICA-E3D (EA 4426)  
amar.lakel@u-bordeaux-montaigne.fr

Résumé : L'information de santé sur internet est un des champs couvert par de nombreuses études dans les web studies. Les enjeux de santé publique et la mutation de l'économie éditoriale numérique en font tout naturellement un objet de recherche, mais aussi un objet de controverse publique. Dans le cadre de cet article, nous nous proposons d'appliquer un nouveau protocole de recherche afin de dégager une sociologie des influenceurs sur internet à travers l'étude du discours sur l'asthme chez les enfants. Cette contribution se veut aussi une discussion/proposition sur les enjeux méthodologiques de l'usage des méthodes digitales dans une approche socio-sémantique à travers les concepts de « prise de position » et d'influence. Comment analyser les positions stratégiques des locuteurs dans un corpus hétérogène et complexe ? Comment cette analyse contribue à l'opérationnalisation d'une sociologie de la controverse sur le web ?

Mots clés : information de santé, Internet, méthodes digitales ; influence, prise de position, socio-sémantique

## **Introduction**

Devenu, au fil des années, l'espace de publication et de conversation préféré des organisations et des personnes, Internet est, aujourd'hui, une bibliothèque hétérogène de documents numériques qui, depuis longtemps, ont fait exploser les seuils de l'archivage tel qu'on l'imaginait à l'époque du papier (pour une réflexion assez complète sur les bouleversements des « macro-systèmes techniques » numériques voir Robert, 2005). Mais loin d'être une aubaine pour les chercheurs et les experts, le web comme espace ouvert et complexe

est un défi pour l'analyse des nouveaux processus d'éditorialisation (Lipsyc et Ihadjadene, 2014 ; Vitali Rosati, 2016 ; Simonnot, 2016). Le web est un paradoxe : gigantesque archive, il est toutefois très difficilement utilisable comme source de connaissance. Face à cette infobésité, les acteurs sociaux, et en premier lieu les professionnels de l'information, ont dû très vite s'organiser pour développer des outils d'intermédiation entre le corpus numérique et les besoins des lecteurs. L'économie numérique a été avant tout une économie d'accès à l'information par une intermédiation algorithmique (moteurs de recherche) et collective (socials bookmarkers et réseaux sociaux)(Le Deuff, 2012).

Le secteur de la santé, mis au défi par cette révolution discursive, est entré dans l'ère du numérique par une double problématisation de la gouvernance de la prescription dérégulée et de l'amélioration de la relation soignant-soigné. La compréhension de l'information de santé numérique comme contexte de la démarche de soin est devenue un impératif dans une politique de santé contemporaine. C'est dans ce contexte que la consultation d'informations de santé sur internet par les usagers est devenue, dès le départ, autant un problème de santé publique qu'une opportunité dans l'appui à l'éducation pour la santé. La prise d'autonomie des usagers vis-à-vis des autorités de santé, constatée par la montée en puissance des associations de patients (Akrich et Méadel, 2002 et 2007), a débouché sur le foisonnement de blogs, de forums ou de sites Internet produits par des patients, mais aussi par des « prescripteurs » non praticiens voire par des médecins indépendants de toute institution de tutelle. Une foule d'acteurs chaque jour plus hétéroclites intervient directement dans l'éducation à la santé. Ces nouveaux corpus numériques produisent, aujourd'hui, une dérégulation et une concurrence en matière de "prescription" en ligne : un tiers s'est insinué entre le soignant et le soigné dans le protocole de soin au risque de transformer profondément l'économie de la prescription. Ce tiers a la particularité d'être peu perceptible par les médecins (alors que 2/3 des patients disent s'informer sur leurs maladies seulement

12% l'avouent à leurs médecins)<sup>1</sup> et son influence est souvent sous-estimée ou pire encore fantasmée.

Face à cela, la question du bouleversement de l'économie informationnelle sur Internet en matière de santé reste un champ d'étude en émergence, notamment dans l'aire francophone. Notre recherche pose la question de la nouvelle économie du sens dans un espace qui n'est pas à proprement parlé dérégulé, mais qui a connu une révolution dans son droit à la parole. L'ouverture et la dérégulation du droit à la parole en ligne a bouleversé la sociologie des prescripteurs. Quelle est la place des professionnels de la santé sur l'information de santé dans l'espace numérique ? Subissent-ils la concurrence de nouveaux entrants plus efficaces et plus influents qu'eux en matière de communication en ligne ? Dans ces circonstances, l'enjeu d'une compréhension de la sociologie des acteurs numériques en information de santé n'est pas seulement un enjeu épistémologique pour le chercheur en sciences sociales mais est un enjeu ontologique qui vise à éclairer les acteurs de la santé en France des risques, mais aussi des opportunités de la communication numérique en matière de santé. Cette problématisation pose la question de notre capacité à comprendre mais aussi à analyser les stratégies éditoriales dans de grande masse de données.

Dans le cadre d'un projet de recherche, il s'est agi de saisir jusqu'à quel point la démarche de consultation d'information de santé sur internet de la part du patient et des accompagnants pouvait être aujourd'hui l'objet de discours et de stratégies discursives nouvelles et hétérogènes. L'enjeu du projet était de cartographier l'écosystème numérique en matière d'information de santé afin d'en dégager les acteurs, comprendre leurs « prise de positions », évaluer leur influence et leur offre en matière de prescription. Dans le même temps, nous avons voulu démontrer la faisabilité d'un protocole d'étude globale de l'écosystème web à travers l'exemple des maladies chroniques chez l'enfant. Pour comprendre la position épistémologique qui est la source de ce projet, nous ferons dans un premier temps une revue de

---

<sup>1</sup> CNOM, *Les conséquences des usages d'internet sur les relations patients-médecins*, Enquête IPSOS, 2010.

la littérature en SIC en matière de recherche sur l'information de santé sur internet. Puis nous décrirons notre méthodologie et notre corpus afin d'éclairer les spécificités de notre approche à la recherche d'une opérationnalisation du concept de "prise de position" sur internet.

## La problématisation de l'information de santé sur internet

À partir d'une recherche sur les mots-clés ("information de santé" AND internet) dans Google Scholar, nous avons constitué une bibliographie francophone de 65 articles consacré à l'analyse du web en matière d'information de santé publiés dans des revues scientifiques sur une période allant de 2001 à 2018. Les revues étaient issues, dans une forte majorité (71,9% des cas) du champ disciplinaire de la Santé publique (Sciences sociales et Santé). Les Sciences Humaines et Sociales sont présentes dans moins d'un tiers du corpus (SIC et Sociologie du Web en général).

### Nouveaux régimes discursifs et nouveaux savoirs du patient en ligne à l'aune du web 2.0

L'analyse de l'information de santé dans les forums et sur les blogs par la sociologie web et les SIC est résolument orienté vers l'étude de la parole du patient. Dans une approche qui privilégie le qualitatif au quantitatif dans l'analyse du discours, les chercheurs se veulent compréhensifs face à ce qu'ils pressentent comme une transformation de l'économie du discours sur internet. La mise en avant d'une parole en décalage avec l'institution médicale supporte l'hypothèse majeure que l'entrée de nouveaux types d'actants dans le discours sur la maladie, qui plus est sur internet, a changé les conditions de production du savoir sur l'information de santé et donc de la relation patient-soignant. Deux centres d'expertise émergent dans ce type de recherche SHS. D'une part, le Centre de Sociologie de l'Innovation (CSI) de Mines ParisTech fondé par Michel Callon et dirigé par Madeleine Akrich (Akrich et Méadel, 2002 et 2007; Callon et Rabeharisoa, 2008; Akrich et al., 2009 et 2014; Méadel, 2009). On

retrouve dans ces articles une inspiration de la sociologie pragmatique (Nachi, 2015) et des théories de la traduction de Callon qui use du champ de l'information de santé comme un terrain probatoire. D'autre part, les chercheurs en SIC en général et du GRESEC en particulier (Groupe de recherche sur les enjeux de la communication de l'Université Grenoble Alpes), produisent l'autre moitié de la littérature scientifique en SHS (Convert et Demailly, 2003, Lafon et Paillart, 2007 ; Romeyer, 2008, 2010 et 2012 ; Staii et al., 2010, Paganelli et al., 2014, Bettaĳa, 2016).

L'ensemble de ces recherches repose sur une approche qualitative se rapprochant du paradigme de l'interactionnisme pragmatique. Si le corpus s'avère ni exhaustif ni représentatif, ces "études exploratoires" reste prudentes quant à la représentativité voire à la généralisation de leur étude du fait de l'impossibilité de constituer ni une population de référence ni un échantillon représentatif. Le parti pris d'une ethnométhodologie numérique sur un échantillon de convenance nous révèle à quel point, aujourd'hui, les méthodes digitales nécessitent la constitution d'outils et de méthodes qui restent encore à inventer (Proulx et Rueff, 2018 ; Cormerais et Lakel, 2019). Une approche consiste à extraire la totalité des posts sur quelques mois d'un forum (Staii et al., 2010), une autre est de comparer deux ou trois blocs d'une liste de diffusion traitant d'un même sujet (Akrich et Méadel, 2002 et 2007 ; Akrich et al., 2009) profitant de la position d'observateur invisible des interactions en place (ce qui soulève une problématique éthique dans la démarche de recherche (Bettaĳa, 2016)). En s'appuyant à la fois sur la description formelle des séquences d'interactions sur le modèle goffmanien et sur une analyse lexicale du discours thématisé, l'approche de ces recherches se centre sur l'émergence des nouvelles stratégies discursives des patients et des aidants.

L'ensemble de ces recherches semble se stabiliser en un champ d'hypothèses réparties sur trois plans : stratégique, pragmatique et stylistique. Sur le plan stratégique, c'est la prise de parole du patient qui semble au cœur de toutes les attentions. Ce dernier serait à la recherche d'une expression de son expérience

personnelle qui lui serait réconfortante et rassurante dans une position discursive spécifique. Il viendrait alors s'informer sur sa maladie en discutant dans un espace communicationnel moins anxiogène que dans le cadre de la relation patient-médecin. Les espaces numériques apparaissent alors comme des espaces performatifs permettant une expression à la fois ouverte et anonyme afin d'aborder un sujet tabou et angoissant. Mais une autre série d'hypothèses s'interroge sur l'empowerment du patient face à l'institution médicale. Avons-nous affaire à une remise en cause des partages du pouvoir, des hiérarchies et des distributions institutionnalisées des rôles ? Dans la même veine, les patients coalisés dans des espaces numériques ne participeraient-ils pas à la construction de leur parcours de soins en élaborant des projets collectifs et des savoirs partagés ? Le modèle de la montée en puissance des associations de patients à la fin des années 80 en France offre un ensemble d'hypothèses transférable au sein des socialisations numériques. En somme, nous assisterions à une mise en ligne de l'opposition traditionnelle entre le modèle hiérarchique de Parsons et le modèle co-constructif de Strauss dans la relation soignant-soigné (Castel, 2005 ; Paganelli et al., 2014). Alors vient, en troisième hypothèse, la question d'une structuration, ou du moins d'une formalisation progressive, d'un savoir expérientiel propre à la communauté de patients qui serait le fondement légitime du patient expert : Un patient capable de co-construire le parcours de soin. Enfin la question de la gouvernance de ces espaces numériques *a priori* libre pose de nombreuses questions. Quid des associations de patients et plus globalement des formes de socialisation des patients hors-ligne c'est-à-dire préexistants, ou du moins existants en dehors, des mondes numériques ? Quid des règles d'engagement et de régulations au cœur même de ces espaces numériques (Vercher et Touboul, 2010) ? Voyons-nous émerger de nouveaux modes d'échanges, de nouveaux rôles qui pourraient être issus de ces espaces numériques ? Dans le cadre de controverse du type mailing-listes ou forums, le recadrage du discours peut passer de la non-réponse et de l'ignorance manifeste au bannissement. On le sait, derrière la revendication normative de la liberté absolue de la parole, les forums ne sont pas exempts d'une gouvernance soft appuyée sur des leaders jouant de la

« coersédution » (Ravault J.R., 1986) pour maintenir un certain ordre du discours (Lakel et Massit Folléa, 2007).

Ce qu'apporte la sociologie du web à la question de l'information de santé sur internet.

Les recherches en sociologie du web et en SIC dans le champ de l'information de santé, alors même qu'elles sont encore en phase exploratoire, permettent aujourd'hui de dégager des pistes de recherche claires sur le rôle d'internet dans la transformation de la relation patient-soignant. L'analyse des associations des patients a montré le rôle très institutionnalisant de ces instances. Elles tirent leur légitimité tant dans la représentation des patients que dans une fonction de diffusion des politiques de santé au point parfois de devenir les relais de communication de la parole institutionnelle (Ackrich, 2002). Dans cette position d'intermédiation, trois fonctions communicationnelles se dégagent alors : traduire-médiatiser, évaluer-diffuser les bonnes pratiques, participer-co-construire le savoir en tant que patient actif. Au regard de cette stratégie communicationnelle, les espaces numériques n'apparaissent pas comme des formes structurées d'actions collectives (Ackrich, 2002 et 2007), mais comme des espaces collaboratifs ouverts et libres (Convert et Demailly, 2003). Si la médecine est très peu critiquée voire très vivement défendue quand elle est concurrencée par des voies "alternatives", le médecin peut voir, quant à lui, sa compétence individuelle remise en cause (Convert et Démailly, 2003 ; Staii et al., 2010; Paganelli et al. 2014). Le consensus sur la science médicale n'est pas nécessairement un consensus sur l'exercice d'un médecin particulier que l'on veut évaluer. Émerge alors une forme d'engagement nouveau que l'on pourrait qualifier d'individualisme collectif autorégulé (Convert et Demailly, 2003) où chaque acteur reste seul maître de ses choix, mais conseillé, voire motivé par le collectif, acceptant très rarement de s'agréger à un collectif de représentation (sondage, pétition, etc.) (Ackrich, 2007) ou dans la construction d'un savoir partagé (portail, logiciel, bibliographie ou webographie, etc.). L'espace numérique reste un espace public ouvert et sans engagement contraignant. Dans le dialogue avec les institutions (publiques ou privées) de santé, il se

situe à mi-chemin entre les collectifs institués et l'individu.

Ainsi, loin d'être des espaces "révolutionnaires" contestant un quelconque pouvoir médical, les espaces numériques étudiés sont véritablement des "forums ouverts et complexes" (Callon et ali., 2014) de délibération et de discussion du savoir médical. Il joue ainsi une fonction de médiation dans une grande proximité avec le patient. Cette médiation est donc tout à la fois un travail d'explicitation des mondes médicaux qu'une évaluation des champs d'expérimentation des savoirs profanes. En cela, leur fonction est très proche de celui d'une association de patients comme caisse de résonance du diagnostic, mais l'institution en moins. Très souvent la thérapie ou le médicament nécessitent des explicitations et une traduction en des termes plus courants. Explication que le colloque traditionnel du soignant n'a pu ou n'a su fournir. Peuvent circuler dans ces espaces des discussions et des propositions sur des accompagnements thérapeutiques qui sont hors du champ de la médecine et du médecin. S'ils ne sont jamais proposés comme des alternatives, elles sont bien souvent sur des registres qui échappent aux champs de la biologie. Homéopathie, méditation, hypnose, régime, herbothérapie de confort sont des savoirs expérimentaux qui se discutent sur les forums en ligne et ce à l'initiative des patients.

Mais une des nouvelles problématiques qui émerge dans ce champ de recherche, c'est l'émergence d'un nouveau régime discursif très présent dans ces espaces numériques et qui leur donne une spécificité propre. On sait depuis longtemps que la question du bien-être du patient dans sa vie quotidienne n'est pas affaire uniquement de médecine. Pourtant, les sujets qui traitent de l'impact de la maladie sur la vie quotidienne forment plus de la moitié des sujets de discussion en ligne. C'est un régime du témoignage personnel, de l'expérience vécue à titre individuel qui fonde cette prise de parole. La volonté de reconnaître la spécificité et les besoins de chaque cas comme à la fois unique et partagé (Battaïa, 2013 ; Paganelli et al., 2014) autorise au témoignage personnel. Les récits de vie forment une information hybride entre privé et scientifique. Information hybride qui mélange discours objectivant et plainte subjectivante, volonté de savoir et



expression affective, recherche de solution et reconnaissance de la souffrance et de l'angoisse (Battaïa, 2010 et 2013). En contrepoint du discours médical, souffrance corporelle et angoisse psychique retrouvent leur unité dans un même régime du pathos. Dans ce régime "l'information est émotionnelle et l'émotion est informative" (Staii, et al, 2010). Formellement le style semi-oral empreint de fautes d'orthographe ou d'idéogrammes phonétiques est revendiqué comme une liberté de parole personnelle. L'anonymat est d'ailleurs un droit fondamental associé au régime du témoignage et tout un jeu d'intégration progressive est lié au dévoilement de son identité (Convert et Démailly, 2003). Pour autant si le droit à communiquer, fondé sur une égalité et une symétrie totale, est essentiel dans les espaces publics numériques, le discours respecte des règles invisibles qui régulent les communautés en ligne selon une nétiquette pas toujours formalisée. Par l'exercice de la parole, se constitue une hiérarchie de fait non explicite qui va du "lurker" silencieux à "l'expert" en passant par l'amateur naïf et l'engagé (Paganelli et al., 2014) révélant une autorégulation active.

## Mesurer l'impact des stratégies d'acteurs sur le web en information de santé

Notre protocole de recherche s'est, dès l'origine, positionné comme complémentaire de ces deux approches. Un outil d'extraction (*crawler*), d'archivage et de qualification du Web (My Web Intelligence<sup>2</sup>.) nous a aidé à constituer un corpus numérique ouvert pour la recherche en sciences sociales. « Il faudrait pouvoir tout lire » nous avait enjoint Michel Foucault<sup>3</sup>. Notre approche a consisté, dans cette logique, à étudier le corpus le plus large possible sans aucun tri a priori afin de fournir une étude de population non filtré en s'appuyant sur une grande masse de données.

---

<sup>2</sup>My Web Intelligence est aujourd'hui un prototype de plateforme créé au sein du MICA et qui tourne en environnement python sur base de données SQLite : <http://mywebintelligence.net>

<sup>3</sup> Michel Foucault, « Michel Foucault, Les Mots et les Choses » (entretien avec R. Bellour), Les Lettres françaises, n° 1125, 31 mars-6 avril 1966, pp. 3-4 : « On devrait tout lire, tout étudier. Autrement dit, il faut avoir à sa disposition l'archive générale d'une époque à un moment donné. »

À partir de la requête “asthme (enfant OR adolescent OR jeune OR bébé OR nourrisson OR infantile)” soumise au moteur de recherche de Google, nous avons extrait 1235 pages de premier niveau (*depth=0*) que nous avons groupé en un second set de données de 846 domaines d’expression<sup>4</sup>. Dans cette recherche sur la problématisation des influenceurs en information de santé sur internet, nous sommes partis sur un scénario d’usage d’un internaute français lambda s’interrogeant sur l’asthme des enfants dans Google (sujet qui compte plus de 90% des requêtes en France<sup>5</sup>). Les termes ont été choisis volontairement communs et courts pour embrasser un échantillon de corpus le plus ouvert possible. Nous avons voulu éviter dans cette recherche le biais d’échantillonnage introduit dans les études précédentes par un vocabulaire socialement situé<sup>6</sup>. Nous l’avons vu, le choix d’un vocabulaire situé socialement surreprésente socio-sémantiquement les populations utilisant habituellement ces mots-clés.

Au niveau de la page, une première extraction automatique obtenait le code *HTML*<sup>7</sup> complet, le titre (*title*), l’*url*, la *description* contenue dans les balises méta. Puis des algorithmes assurent une première couche d’extraction en détectant le contenu principal du document (*voir note 10*) ; les URL de liens (*urls*), d’images (*imgs*) et de vidéos (*videos*) au cœur de ce contenu. En cas d’extraction par crawling à plusieurs niveaux de profondeur, la variable *depth* nous rappelle si le document est arrivé au démarrage (niveau 0) ou par extraction (niveau n). À ce premier niveau d’extraction s’ajoute un niveau automatique d’extraction externe via des services web par api. On trouve à ce niveau la résonance sociale de l’URL (*facebook\_like*,

---

<sup>4</sup> Le domaine d’expression n’est pas tout à fait le nom de domaine car dans le cas des fermes de contenu (réseaux sociaux, plateforme de blogging, etc.) nous avons dû développer des heuristiques de découpage de ces noms de domaines en sous domaine correspondant aux unités éditoriales (pages, profils, groupes, etc.)

<sup>5</sup> Pour des données actualisée voir : <http://gs.statcounter.com/search-engine-market-share/all/france/#monthly-201803-201903-bar>

<sup>6</sup> Nous avons aussi utilisé la fonction de navigation anonyme de chrome et de Firefox pour supprimer les biais d’historicité et d’identification avec un Proxy VPN anonymisant situé à Paris (ExpressVPN) pour éliminer le biais de localisation. Les données ont été récoltées entre le 15 mars 2016 et le 21 mars 2016.

<sup>7</sup> Par convention nous mettrons les noms de variables en italique tout au long de cet article

*facebook\_share, linkedin\_share*)<sup>8</sup>. Dans un troisième temps, une annotation humaine va permettre de contextualiser socialement le document grâce à trois variables : le type de média (*media\_type*), le type d'émetteur (*emitter\_type*) que nous avons voulu volontairement ouvert.

## Les acteurs de l'information de santé sur le web en France

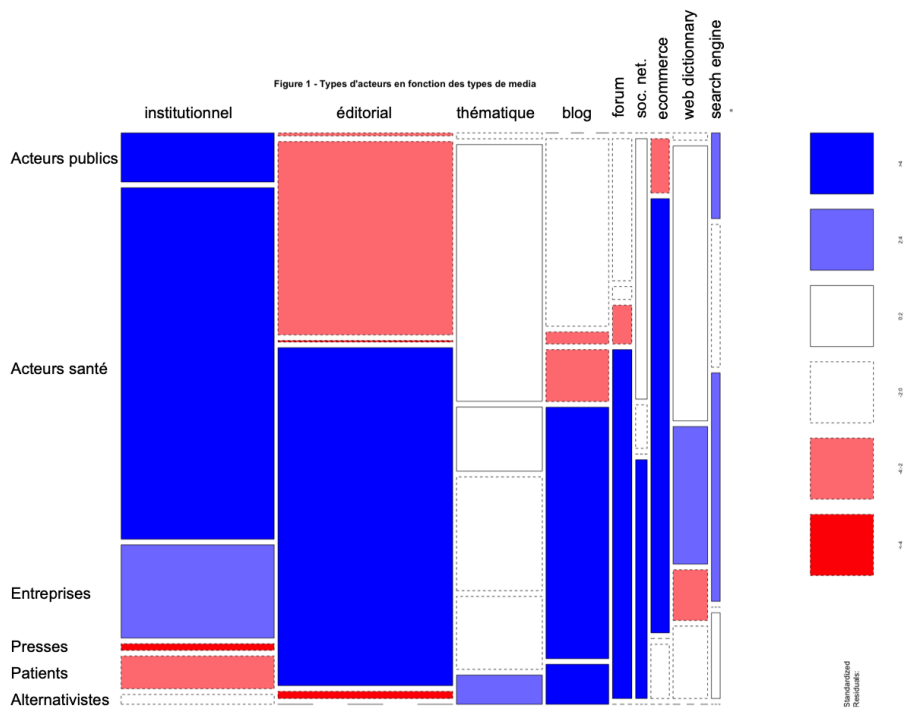
Dans l'étude de la typologie des locuteurs, on constate que le discours est dominé par les acteurs de santé et par les journalistes (60,8% des sites et 67,9% des pages). Les patients et les entreprises forment le troisième tiers des acteurs de ce champ (avec respectivement 13,6% et 13,3% des pages). Les acteurs publics hors santé (3,1% des pages) et les alternativistes<sup>9</sup> (2,10% des pages) jouent, quant à eux, des rôles très minoritaires. Les plateformes communicationnelles les plus utilisées sont les sites éditoriaux avec une forte stratégie de contenus périodiques (31,8 % des sites), les sites de présentation institutionnels (25,8 % des sites) et les plateformes du web 2.0 que sont les réseaux sociaux, les forums et les plateformes de blogs (17%). Viennent ensuite les sites de ressources type annuaires (7,3 % des sites) et une faible représentation des sites directement e-commerce (3,9%). L'utilisation du X2 a permis de comparer les répartitions des acteurs pour constater leurs différences sur et sous-représentations en fonction de leur choix de plateforme. Apparaissent alors des mondes très distincts en matière de « lieux » de communication. Le monde de la communication institutionnelle attire très fortement les acteurs de santé et les institutions publiques qui ont en commun une très forte répulsion des stratégies de contenus et des réseaux sociaux. Les journalistes et les patients sont les plus éloignés de cette stratégie « corporate », les premiers préférant très exclusivement la stratégie éditoriale. Ce serait une lapalissade de dire que les journalistes préfèrent les sites éditoriaux si ce n'est de

---

<sup>8</sup> Il faut souligner à ce stade que de nombreux infomédiaires ont développé un travail de qualification d'urls qui sont autant de source d'enrichissement d'analyse du web. Dans notre cas <https://www.sharedcount.com/api-docs>

<sup>9</sup> Nous classons dans cette catégorie tous les acteurs qui proposent des thérapies plus ou moins globales basées sur des méthodes non scientifiques. Principalement des méthodes basées sur l'usage d'éléments naturels, magiques ou ésotériques.

constater qu'ils sont les seuls à posséder une attraction pour cette stratégie de contenu et sont les principaux producteurs de contenus périodiques et actualisés. Les patients (avec les alternativistes) préfèrent radicalement une stratégie conversationnelle social media faite d'interventions dans les forums et pour les plus engagés dans des plateformes de blogs. Les entreprises quant à elles s'investissent pleinement dans les sites de e-commerce, de ressources et institutionnels tout en négligeant en apparence les sites de contenu et social media.



Source : Auteur

Ainsi face à l'hypothèse d'un bouleversement de l'espace public médiatisé par le web, il nous faut constater que, dans le champ de l'information de santé, les acteurs occupent des positions très distinctives, mais qui restent dans un ordre communicationnel traditionnel. Dans l'espace public numérique médiatisé par Google, la

vague de la parole des patients et des associations de patients est très minoritaire face aux professionnels de santé. Ces derniers ont su sortir de leur réserve institutionnelle quand ils ont pris une forme associative ou à titre individuel. L'innovation communicationnelle en matière d'information de santé semblant inversement proportionnelle à son institutionnalisation. Les alternativistes sont, quantitativement, une pincée négligeable dans la prise de parole et n'apparaissent pas susceptible de remettre en cause l'ordre établi. Ce qui en revanche frappe très clairement, c'est la puissante présence de l'industrie journalistique à travers la presse générale, spécialisée et médicale qui occupe la première position en termes de prise de parole dans le champ de l'information de santé.

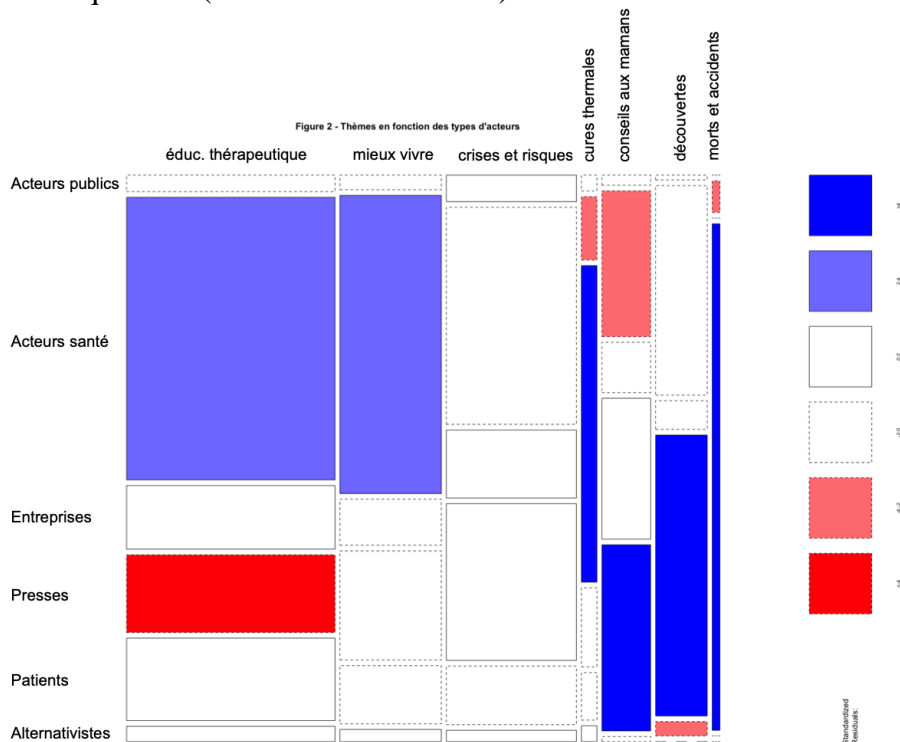
Nous avons ensuite classé les prises de parole des messages par une méthode de text mining en catégorisant automatiquement les *contenus* selon une position vectorielle à  $n$  dimensions grâce à la librairie *tm* de R. La méthode consiste à réduire le contenu, préalablement « readabilisé »<sup>10</sup>, en une série de mots-clés lemmatisés qui sert de base à un tableau disjonctif à 790 dimensions (le dictionnaire global des lemmes retenus) qui vient donner l'empreinte numérique de chaque contenu mis en ligne<sup>11</sup>. *My Web intelligence* est capable d'isoler le contenu éditorial principal de n'importe quel page de son environnement contextuel (menu, barres latérales, pied de page, etc) : son *readable* dans le jargon du web. Dans le logiciel R, nous avons utilisé la library de text mining *tm* pour réduire le texte à un dictionnaire signifiant en enlevant les mots courants (« stop words ») et en lemmatisant le vocabulaire restant. Ainsi chaque expression est réduite à un sous-dictionnaire qui le positionne dans un monde sémantique à  $n$  dimensions formé par la totalité du corpus. Chaque expression sera qualifiée dans une matrice constituée d'un dictionnaire global pertinent à l'échelle du corpus (790 stems dans

---

<sup>10</sup> le processus s'appuie essentiellement sur la library Beautiful Soup de Python et consiste à supprimer l'ensemble du code HTML périphérique pour ne garder que le contenu central du document en format texte. La procédure nécessite encore un nettoyage manuel à posteriori pour ne garder que le texte du locuteur central de la page.

<sup>11</sup> Cette catégorisation est très proche de la méthode utilisée par le logiciel Iramuteq et Alceste mais nous avons préféré la library *tm* dans R qui nous a offert un meilleur contrôle des opérations par la suite.

notre cas). Il nous faut ensuite réduire ces 790 dimensions à 5 par l'usage d'un ACP et finaliser la catégorisation par une méthode de clustérisation. Les clusters que nous appellerons thèmes sont produits par une classification ascendante hiérarchique (CAH) renforcée par la méthode des K-Mean. 7 thèmes majeurs ont été discriminés statistiquement (les *thèmes* du discours)<sup>12</sup>.



Source : Auteur

Le thème principal qui forme le cœur des représentations vise à informer sur les caractéristiques de la maladie et les comportements à adopter dans une approche d'éducation thérapeutique (64% des pages). Il se distingue en 3 axes thématiques. Le premier concerne la description de l'asthme chez l'enfant, ses symptômes, ses origines et

<sup>12</sup> Cette méthode a l'avantage d'être entièrement automatisable et donc échelonnable dans le traitement de grand corpus. Pour une meilleure lecture nous avons fait un travail de recodage manuel pour apporter un sens courant à cette classification automatique

ses conséquences (37% des pages), ainsi qu'un discours d'accompagnement et de conseil pour les mamans sur la façon de mieux vivre avec l'asthme de l'enfant (18,3% des pages) et plus précisément sur les bons gestes à faire pour le bébé et le nourrisson (8,7% des pages). Un deuxième axe thématique vient porter un discours à la charge émotionnelle très forte autour des facteurs provoquant des crises graves et des risques sérieux encourus par l'enfant asthmatique (23% des pages). S'ajoute à cela un troisième axe plus sensationnaliste qui occupe 10,5% des positions avec des thèmes aussi variés que la mort d'un nourrisson atteint d'asthme dans les urgences d'un hôpital (une polémique récurrente qui a occupé 1,4% des prises de parole de notre corpus) que les études sensationnelles sur la pollution automobile, les bactéries intestinales, la présence d'animaux domestiques (le chien en particulier) et l'interaction médicamenteuse. Vient enfin la promotion de cures thermales pour les asthmatiques (2,7%) par des entreprises dans le secteur de la santé. Ainsi près de 80% des pages étudiées sont directement liées à une volonté d'éducation du patient plus ou moins alarmiste. Une infime minorité est constituée de pages sensationnelles qui tentent, par le scoop, de faire le buzz.

### Opérationnalisation et formalisation le concept de "prise de position" sur le web

Devant la forte corrélation des variables *emmitter\_type*, *media\_type* et *themes*, nous avons opérationnalisé la notion de *position stratégique* par la construction d'une variable éponyme à partir de ces trois variables. L'usage d'une nouvelle CAH renforcée a permis de dégager huit clusters [numéros entre crochets] en sept « positions stratégiques » dans le corpus étudié. Cette nouvelle variable a été analysée en opérant une série de tables de X2 entre cette "position" et les 3 variables constitutives pour interpréter les qualités constitutives de ces positions stratégiques du discours :

- La plus importante stratégie [7], qui occupe 21,3% des pages, surreprésente très fortement les institutions de santé (resid:16.44), les institutions publiques (8.56), les associations de praticiens et

les universités (respectivement : 6.57 et 6.14) et enfin les sites d'assurance et de mutuelles (4.17). Le support médiatique favori est de très loin le site institutionnel (resid:21) abordant les thèmes d'éducation thérapeutique (resid:4.12) et de conseils pour mieux vivre son asthme (résid:2.57). Nous sommes là en présence de la position institutionnelle en matière de communication de santé.

- La seconde position [2] (18,6% des discours) est celle de la presse médicale (13.14) et de la presse spécialisée (12.9) qui utilisent prioritairement des médias éditoriaux (17.5) pour aborder des thèmes à forte valeur émotionnelle tels que les facteurs de crises graves (3.04) et les découvertes sensationnelles (8.05). Cette presse spécialisée s'engage dans le processus d'éducation thérapeutique, mais par le filtre de la nouveauté et de l'exceptionnel.
- Il faut ajouter à cette position, une position très proche (concurrentielle ?), mais plus spécialisée sur le thème de l'asthme [4]. Cette troisième position (13,7% des pages), constituée principalement d'associations de praticiens (6.64), mais aussi d'une petite part de la presse médicale et spécialisée (2.57 et 2.39), s'attache à être le relais de l'information de santé tant en matière de mieux vivre son asthme au quotidien (3.92) que de prévenir les risques et les crises graves (2.42). Cette position est située à mi-chemin entre celle des institutions de santé et celle de la presse médicale et spécialisée. Moins enclins au sensationnalisme que la position précédente, le poids des praticiens en associations permet tout de même de couvrir une information plus grand public liée aux préoccupations du patient et des accompagnants.
- À l'inverse la quatrième position [1] (10,3% des pages) est celle par laquelle le scandale est mis sur la place publique. Constituée essentiellement par la presse généraliste (30.72) utilisant principalement des sites éditoriaux (13.58), ces pages sont clairement à la recherche du scandale et de l'accident (11.53) et dans une moindre mesure des découvertes sensationnelles (4.51) qui vont faire de l'audience et faire vendre. On retrouve ici l'archétype de la presse sensationnaliste portée par les presses



quotidiennes.

- La cinquième position [5] (14% des discours) représente la parole individuelle structurée et engagée. On y retrouve le discours du patient (14.87) au côté du praticien individuel (13.02) et d'une partie des alternativistes (3.33). Ils investissent très fortement le web 2.0 avec une très forte préférence pour les blogs (22.55) et les réseaux sociaux (13.02). Leur thématique préférée reste l'éducation thérapeutique qu'ils disputent aux institutions de santé en abordant des pratiques laissées à la marge du corps médical. Cette catégorie constitue l'avant-garde active du web 2.0 tout aussi experte que désinstitutionnalisée.
- La 6ème position [3] (9% des pages) est très proche de la précédente en ce qu'elle engage principalement des associations de patients (11.08), des patients à titre individuel (4.11) et de la presse spécialisée en moindre mesure (2.53), mais cette fois-ci principalement dans les forums (19.67). On retrouve ici le lieu de discours très étudié dans la sociologie du web où le patient et l'accompagnant sont à la recherche de conseils pratiques dans l'accompagnement du bébé et du nourrisson avec sa maladie (22.24).
- Enfin la 7ème et dernière position [6 et 8] (13,11% des pages) est celle de la stratégie commerciale des entreprises (22.23) qui utilise une gamme de médias très complète sur la base du site e-commerce (25.65), mais aussi de portails et d'annuaires (6.35 et 4.46), de sites institutionnels (3.10) et thématiques (2.44) afin de vendre des cures thermales et des produits de soins afférents (crème et huiles, etc.). On voit se déployer alors une stratégie de communication très complète sur la base d'une stratégie institutionnelle, de référencement et d'influence.

Si la parole des patients (5ème et 6ème position) a été le champ privilégié de la recherche en SHS et que la parole des professionnels de santé (1ère et 3ème position) a fait l'objet d'évaluation continue dans la recherche en santé publique, force est de constater que les autres positions, qui jouent un rôle majeur dans la construction des représentations, a été peu étudié dans la recherche francophone. Notre

méthode globale, en minimisant les a priori de constitution de corpus, a d'ores et déjà révélé de nombreux angles de recherches peu explorés, démontrant par-là l'utilité d'une recherche préalable à large spectre avant d'appréhender un champ émergent dans des analyses plus micro. Le rôle de la presse dans l'information de santé sur Internet est paradoxalement un angle mort de la recherche alors qu'elle avait été un enjeu majeur dans la recherche sur les médias traditionnels.

## Conclusion

En partant d'un scénario d'information banal, nous avons rencontré les limites d'une constitution de corpus cadrée par un moteur de recherche. En donnant accès à un petit millier d'URL par des biais algorithmiques secrets et complexes, Google nous donne sa vision du web (Cardon, 2011 et 2013). Pourtant, le monde selon Google est bien plus diversifié que d'aucuns ont pu le critiquer. Il suffit de voir la répartition en loi de puissance de la résonance sociale, du nombre de liens entrants, mais aussi de l'audience moyenne des sites éditeurs pour supposer raisonnablement que l'échantillonnage fait par le moteur recherche a au moins une homologie structurale avec son index général. Elle semble déjà loin la réflexion sur le Page Rank (Blin et Pages, 1999) comme infomédiaire (discuté par Cardon, 2013), on dénombre aujourd'hui pas moins de 200 paramètres à la construction du fameux ranking (Dean, 2013). Sachant que cet algorithme est secret et complexe, sa rationalité semble tout simplement avoir échappé à ses inventeurs pour devenir un champ de recherche à part entière. Devant l'impossibilité d'appréhender le fond numérique et dynamique, quasi « vivant », du web, il nous semble pertinent, dans le cadre du web studies, de cadrer sa recherche à des mondes numériques (moteurs de recherches, réseaux sociaux, applications mobiles, serveurs de mails, etc) liés à des scénarios d'usage et d'expositions informationnelles.

Dans le cadre de notre recherche, nous avons voulu inclure l'analyse des citations explicites comme indicateur de légitimité (Cardon et al., 2014). Pour autant, notre méthode avait voulu rompre avec les cartographies du web traditionnelles qui absorbent la totalité

des liens d'une page (ex. les outils de Hyphe du medialab ou Linkfluence). Pour qui a l'habitude des pages web modernes, les liens d'une page ne sont pas assimilables à des citations explicites tant elles sont hétérogènes dans leurs sources, leur production et leurs natures. Il nous a semblé qu'une analyse sociosémantique à partir de ces liens était une contradiction en soi. Nous avons alors fixé deux contraintes à l'analyse du lien. Qu'il soit une citation dans le corps du texte et qu'il mène à une page qui soit un individu de notre territoire sémantique. Pour cela nous avons recours à l'analyse structurale des réseaux comme méthode d'analyse des pages et des sites web (Wasseman et Faust, 1994 ; Scot, 2012). Mais il a fallu très vite faire le constat d'une extrême faible connectivité des pages. Seul 1,5% des pages ont plus d'un lien qualifié entrant (indegree). La limite de notre construction de corpus, limité au web rendu visible par Google dans un scénario d'usage et à des liens sémantiquement qualifiés a rendu donc inopérant cette recherche du moins au niveau du réseau des pages. Nous envisageons, par la suite, de mener des études à partir d'un véritable échantillonnage boule de neige issue d'un crawl de proche en proche capable de sélectionner des documents pertinents au regard d'un dictionnaire apriori. Seule une recherche approfondie sur de grands corpus permettrait alors d'étudier la citation comme indicateur de légitimité.

Pour autant, notre recherche a montré que les positions stratégiques des énonciateurs sont suffisamment typées pour qu'on puisse qualifier les microcosmes numériques découpant l'espace public numérique en champs socionumériques. La presse et les professionnels de santé restent les acteurs dominant en information de santé sur internet avec plus de 70% des pages produites. Face à eux la parole du patient est très minoritaire (14%). Dans le monde selon Google, il n'y a donc pas de disruption de la parole institutionnelle en matière d'information de santé. Même si la presse est constamment à la recherche d'informations sensationnelles, de nouveautés et de faits hors normes dans une logique de breaking news et d'audimat, il existe à la fois une spécialisation et une hiérarchisation dans les stratégies d'acteurs qui partent de l'éditorialiste de l'information sensationnelle et exceptionnelle à l'éducation thérapeutique médiatisée par des professionnels de santé. Le cœur de réseau en termes d'influence reste

un jeu à trois hiérarchisé entre presse médicale, associations de praticiens et institution de santé. La discussion des soucis intimes et quotidiens du patient et la part du business de la santé (en l'occurrence ici les cures thermales) restent très faibles dans l'espace public numérique étudié. Si l'on ajoute la très faible présence des alternativistes (moins de 3% des pages accessibles dans les moteurs de recherche) dans le corpus, force est de constater que l'ordre social du discours en matière d'information de santé est très peu remis en cause à l'ère numérique.

## Bibliographie

Akrich, M., & Meadel, C. (2002), « Prendre ses médicaments/prendre la parole : les usages des médicaments par les patients dans les listes de discussion électroniques », *Sciences sociales et santé*, 20(1), 89-116. <https://doi.org/10.3406/sosan.2002.1546>

Akrich, M., & Meadel, C. (2007), « De l'interaction à l'engagement : les collectifs électroniques, nouveaux militants de la santé », *Hermès, La Revue*, 47, 145-153.

Akrich, M., Meadel, C., Rémy, C., & Vergnaud, F. (2009). *Les patients et l'information : le cancer au risque d'internet*. Rapport de recherche.

Akrich, M., Moreira, T., & Rabeharisoa, V. (2014), « Evidence-based activism: Patients', users' and activists' groups in knowledge society ». *BioSocieties*, 9(2), 111-128. <https://doi.org/10.1057/biosoc.2014.2>

Battaïa, C. (2013). *L'émotion dans les forums de discussion : structuration et évaluation de l'information de santé*. Consulté à l'adresse <http://www.theses.fr/2013GRENL012>

Battaïa, C. (2016), « Information médicale et émotion dans les forums de santé » *Les Cahiers du numérique*, 12(1), 51-72.

Brin, S., & Page, L. (1998), « The anatomy of a large-scale hypertextual Web search engine », *Computer Networks and ISDN Systems*, 30(1-7), 107-117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)

Callon, M., & Rabeharisoa, V. (2008). « The Growing Engagement of Emergent Concerned Groups in Political and Economic Life: Lessons from the French Association of Neuromuscular Disease Patients », *Science, Technology, & Human Values*, 33(2), 230-261. <https://doi.org/10.1177/0162243907311264>

Callon, M., Lascoumes, P., & Barthe, Y. (2014). *Agir dans un monde incertain. Essai sur la démocratie*. Paris: Points.

Cardon, D. (2013). « Dans l'esprit du PageRank », *Réseaux*, 177, 63-95.

Cardon, D., Fouetillou, G., & Roth, C. (2014). « Topographie de la renommée en ligne », *Réseaux*, 188, 85 -120. <https://doi.org/10.3917/res.188.0085>

Castel, P. (2005), « Le médecin, son patient et ses pairs », *Revue française de sociologie*, 46(3), 443-467. <https://doi.org/10.3917/rfs.463.0443>

Convert, B., & Demailly, L. (2003), « Internet et les professions de sante », *Réseaux*, 120(4), 241-269.

Convert, B., & Demailly, L. (2003). « Les interactions dans Internet », *Communication. Information médias théories pratiques*, (Vol. 22/2), 63-98.

Cormerais F. & Lakel A.(2019), « Recherches digitales et production des données, bouleversement des agencements pour le chercheur en SIC », *Etudes Digitales*, 6, Garnier classiques.

Dean, B. (2013). Google's 200 Ranking Factors: The Complete List (2019). Consulté 28 mai 2019, à l'adresse Backlinko website: <https://backlinko.com/google-ranking-factors>

Lafon, B., & Paillart, I. (2007), « Malades et maladies dans l'espace public », *Questions de communication*, 11, 7-15.

Lakel, A. (2014). My Web Intelligence : Un projet d'outils open source au service des humanités numériques. [online] My Web Intelligence. Available at: <http://mywebintelligence.net/outils-open-source-humanites-numeriques/> [Accessed 16 Nov. 2014].

Lakel, A., & Le Deuff, O. (2017). « À quoi peut bien servir l'analyse du web ? », *Les Cahiers du numérique*, 13 (3), 39-62.

Le Deuff, O. (2012). *Du Tag au Like*. FYP Éditions.

Lipsyc, C., & Ihadjadene, M. (2014), « Architecture de l'information et éditorialisation », *Études de communication*, n° 41(2), 103-118.

Meadel, C. (2009), « Le spectre « psy » réordonné par des parents d'enfant autiste », *Politix*, (73), 57-82.

Nachi, M. (2015). *Introduction à la sociologie pragmatique*. Armand Colin.

Paganelli, C., Clavier, V., & Duetto, A. (2014). « S'informer via des médias sociaux de santé : quelle place pour les experts ? » *Le Temps des médias*, (23), 141-143.

Proulx, S., & Rueff, J. (2018). *Actualité des méthodes de recherche en sciences sociales sur les pratiques informationnelles*. Consulté à l'adresse [http://epe.lac-bac.gc.ca/100/200/300/univ\\_laval/centre\\_etudes\\_medias/CEM-ActualiteMethode.pdf](http://epe.lac-bac.gc.ca/100/200/300/univ_laval/centre_etudes_medias/CEM-ActualiteMethode.pdf)

Robert, P. (2005), *La logique politique des technologies de l'information et de la communication : critique de la logique du glissement de la prérogative politique*, PUB, Bordeaux.

Romeyer, H. (2008), « TIC et santé : entre information médicale et information de santé », *tic&société*, (Vol. 2, n° 1).  
<https://doi.org/10.4000/ticetsociete.365>

Romeyer, H. (2012), « La santé en ligne », *Communication. Information médias théories pratiques*, (Vol. 30/1).  
<https://doi.org/10.4000/communication.2915>

Romeyer, H. (2015), « Le bien-être en normes : les programmes nationaux nutrition santé », *Questions de communication*, (27), 41-61.

Simonnot, B. (2016), « Place des moteurs de recherche dans l'éditorialisation du Web », *Communication & langages*, (188), 45-59.  
<https://doi.org/10.4074/S0336150016012035>

Staii, A., Paganelli, C., Romeyer, H., Clavier, V., Gallo, M. C. M., & Mounier, E. (2010). *Dynamiques interactionnelles et rapports à l'information dans les forums de discussion médical*. Consulté à l'adresse <http://hal.univ-grenoble-alpes.fr/hal-01135102>

Vitali Rosati, M. (2016), « Qu'est-ce que l'éditorialisation? », *Sens public*. <https://doi.org/10.7202/1043383ar>

Vercher, E., & Touboul, A. (2010). *Diffusion de l'information et médiation dans le domaine de la santé : nouvelles pratiques, nouvelles règles*. 9.