



**HAL**  
open science

## Information de santé sur internet : le cas de l'asthme et du diabète chez l'enfant.

Amar Lakel

### ► To cite this version:

Amar Lakel. Information de santé sur internet : le cas de l'asthme et du diabète chez l'enfant. : Rapport de recherche "Informations de santé sur internet" Projet Région Innovation numérique des données. Laboratoire MICA. 2018. hal-04217514

**HAL Id: hal-04217514**

**<https://hal.science/hal-04217514>**

Submitted on 25 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Amar LAKEL

Université Bordeaux Montaigne

Laboratoire MICA

[amar.lakel@u-bordeaux-montaigne.fr](mailto:amar.lakel@u-bordeaux-montaigne.fr)

# My Web intelligence

Rapport de recherche

Projet Région Innovation numérique des  
données

2014-2017

Juin 2018

# Sommaire

|   |           |
|---|-----------|
| <b>Introduction</b>   | <b>3</b>  |
| <b>Contexte Général et Epistémologique</b>  | <b>3</b>  |
| Périmètre du projet   | 4         |
| Plan du rapport   | 6         |
| <b>I – La problématisation de l’information de santé sur internet</b>   | <b>7</b>  |
| I.1 - Recherches en santé publique, l’évaluation de la qualité de l’information à l’aune des pratiques médicales                          | 8         |
| I. 2 - Sociologie de la communication : A la recherche des nouveaux régimes discursifs du patient en ligne.                               | 12        |
| <b>I.3 - Ce qu’apporte la sociologie du web à la question de l’information de santé sur internet.</b>                                     | <b>15</b> |
| <b>II – Un dispositif d’intelligence numérique : enjeux méthodologiques</b>   | <b>19</b> |
| II.1 – My Web Intelligence : enjeux des études numériques   | 20        |
| II.2 – De l’enquête des controverses en études digitales : enjeux des digital methods   | 23        |
| <b>II.3 – L’analyse comme re-fabrication du contexte stratégique de l’acteur réseau</b>   | <b>26</b> |
| <b>III – Etude comparée sur les influenceurs en information de santé sur l’internet : le cas du diabète et de l’asthme chez l’enfant.</b> | <b>32</b> |
| III.1 - Quantification du corpus des énoncés numériques   | 33        |
| III.2 - Opérationnalisation et formalisation du concept de “prise de position” sur le web   | 37        |
| III.3 - Mesurer l’influence des acteurs : de la résonance sur les réseaux sociaux à l’authority sur le web                                | 40        |
| <b>Conclusion générale</b>  | <b>45</b> |
| <b>Bibliographie</b>  | <b>48</b> |

# Introduction

## Contexte général et épistémologique

L'avènement de la numérisation a instauré une révolution informationnelle sans précédent, entraînant une prolifération complexe de canaux de communication dans le domaine de la santé publique. Cette diffusion dérégulée et hétérogène d'informations de santé sur des plateformes en ligne engendre une série de dilemmes sociaux, éthiques et politiques, affectant profondément l'écosystème de la prescription médicale. Internet se présente ainsi comme un champ informationnel perturbateur, à l'intersection de multiples cadres épistémologiques, du soignant au soigné, et déstabilise les dynamiques établies depuis des siècles de l'autorité médicale. Par exemple, des mouvements sociaux tels que la vague anti-vaccination ou les discours sur la compatibilité entre médicaments et pratiques religieuses interpellent régulièrement la conduite des politiques de santé publique. Il est évident que des tiers auteurs s'interposent, souvent discrètement, entre le praticien et le patient, modifiant les paradigmes du pouvoir et de la connaissance au sein de la chaîne de soins. Bien que 66,6 % des patients cherchent à s'informer sur leur pathologie via des ressources en ligne, seuls 12 % en informent leurs médecins, exacerbant ainsi une disjonction entre pratiques d'auto-information et diagnostic médical officiel. En 2014, cette reconfiguration épistémique de la gouvernance des prescriptions médicales dans un contexte de santé 2.0 a été à la base de notre projet de recherche.

Dans ce paysage complexe et interdisciplinaire, le projet My Web Intelligence vise à analyser cet écosystème numérique. Il met en lumière les différentes catégories d'acteurs, leurs stratégies discursives et les modalités de leur influence sur le processus de prescription. Cette recherche se situe à l'intersection de plusieurs axes d'investigation, issus de l'expertise pluridisciplinaire du laboratoire MICA de l'Université Bordeaux Montaigne. D'une part, les études sur la "littératie digitale en santé" commencent à émerger en France, comme le démontrent des travaux antérieurs du laboratoire (Le Deuff, 2015). Ces recherches s'articulent autour de concepts connexes tels que la health literacy (Osbourne, 2013; Sorensen, 2015), l'information literacy et la digital literacy. D'autre part, le Centre d'Investigation Clinique Pluridisciplinaire (CIC-P) du CHU de Bordeaux apporte une dimension clinique à cette plateforme expérimentale, visant à élaborer une méta-analyse du discours thérapeutique en ligne. L'enjeu majeur de cette initiative réside dans sa capacité à informer les politiques publiques de santé. Dans une société où l'accès à l'information de santé influence directement la qualité des soins et, par extension, les paradigmes de la santé publique, la gestion de cette information devient essentielle. Les données fournies par l'ARCEP illustrent ce constat : 83 % des Français sont connectés à Internet, consacrant en moyenne deux heures par jour à ce média, qui est devenu en 2016 la première source d'information de santé. Cette recherche vise donc à cartographier le paysage discursif de l'information médicale en ligne, tout en évaluant les mécanismes sous-jacents de

sa hiérarchisation et de sa légitimation. Cette démarche s'impose comme une contribution essentielle à la refonte des stratégies de gouvernance en matière de santé, dans un contexte national et global marqué par la digitalisation accélérée des pratiques sociales et médicales.

Dans le contexte actuel de la médecine, les patients adoptent diverses stratégies pour faire face aux dilemmes thérapeutiques. Selon les statistiques, 59% des patients cherchent à confirmer ou à contester les informations médicales reçues lors des consultations, tandis que 30% participent à des groupes de patients. Ce phénomène peut être interprété comme une tentative de délibération avec le diagnostic médical, un effort pour démystifier et authentifier le discours médical dont ils commencent à douter. En ce qui concerne les maladies chroniques ou socialement stigmatisantes, les patients ont tendance à se connecter à des communautés - qu'il s'agisse de patients ou de soignants - dans le but de rompre l'isolement. La maladie est ici considérée non seulement comme un état médical, mais aussi comme une condition de vie ayant des répercussions psychosociales profondes, nécessitant un soutien émotionnel et affectif.

## **Périmètre du projet**

Il est indéniable qu'une préoccupation croissante se manifeste en France pour l'éducation thérapeutique, considérée comme un moyen de préserver le capital santé. Cette tendance est amplifiée par une couverture médiatique extensive, reflétant un désir collectif d'autonomie sanitaire. Cependant, l'essor de l'économie informationnelle numérique déstabilise cette politique d'éducation à la santé. Le coût marginal de production de contenu est presque nul et la portée est mondiale. Cette démocratisation de la parole publique soulève quatre problèmes majeurs :

1. La décentralisation de l'autorité épistémologique : Désormais, n'importe qui peut se présenter comme un émetteur de "connaissance", indépendamment de son expertise réelle.
2. La primauté du référencement : La visibilité en ligne est souvent dictée par des algorithmes de référencement (et ceux qui savent les manipuler) plutôt que par la véracité ou la pertinence du contenu.
3. L'amplification des informations sensationnelles : Les discours qui suscitent des réactions émotionnelles fortes bénéficient de la viralité inhérente aux plateformes numériques.
4. La difficulté de réguler le contenu en ligne : Les tentatives de régulation technologique ou légale se heurtent à la complexité et à l'échelle de la diffusion de l'information.

En résumé, le paysage numérique de l'information en matière de santé est une mosaïque d'acteurs disparates. Ceux-ci incluent non seulement des professionnels de la santé et des collectifs de patients, mais aussi des entités moins scrupuleuses qui exploitent l'anxiété et la vulnérabilité. La multiplicité et la complexité de ces acteurs nécessitent une approche sociologique nuancée pour comprendre

comment les individus naviguent dans cet écosystème en constante évolution, surtout en ce qui concerne leur santé.

Le projet "My Web Intelligence" est un prototype innovant, financé par le Conseil Régional d'Aquitaine dans le cadre d'un appel d'offres consacré à l'utilisation novatrice des données. Ce projet collaboratif implique divers partenaires institutionnels, dont une équipe de recherche du Centre Hospitalier Universitaire (CHU) de Bordeaux, plus précisément du Centre d'Investigation Clinique Pluridisciplinaire (CIC-P). Supervisé par le Professeur Michael Fayon et dirigé par Amar Lakel, ce projet se situe à l'intersection de la santé publique et des sciences de l'information et de la communication (SIC). L'objectif principal de cette initiative est de cartographier, analyser et mesurer l'influence des acteurs numériques en matière d'information de santé. Plus spécifiquement, il vise à déterminer si une co-régulation de l'information de santé par les professionnels et les experts du domaine est envisageable. Le CHU de Bordeaux envisage plusieurs perspectives en participant à ce projet :

1. Coordination de la relation médecin-patient par le biais du numérique : Avec "My Web Intelligence", le praticien étend son écoute du patient à l'environnement en ligne, permettant ainsi une compréhension plus nuancée de l'expérience individuelle de la maladie.
2. Défense de l'épistémologie médicale : Le projet offre une plateforme permettant aux professionnels de santé de défendre des pratiques basées sur des données probantes contre des discours non scientifiques ou fallacieux qui prolifèrent en ligne.
3. Adoption d'une perspective systémique : Le médecin n'est plus seulement un prescripteur, mais devient également un curateur d'information et un partenaire dans le parcours de soin. Cette démarche collaborative encourage l'évaluation, la labellisation et la promotion de la qualité de l'information médicale en ligne.

Ce projet bénéficie d'une subvention de la Région Aquitaine pour l'utilisation innovante de données, et une convention de recherche a été signée entre l'Université de Bordeaux Montaigne et le CHU de Bordeaux. Ce protocole de recherche, piloté par Amar Lakel, vise à instaurer un processus d'intelligence numérique spécialisé dans le domaine de l'information de santé<sup>1</sup>.

La transformation du sens économique dans le contexte numérique soulève des questions complexes et nuancées, à la fois épistémologiques et ontologiques, qui déterminent notre problématique : Ce changement radical dans l'accès à la parole publique a-t-il transformé de manière radicale l'économie de l'information de santé au point de remettre en cause la relation thérapeutique

---

<sup>1</sup> Qu'il nous soit donné, ici, l'occasion de remercier le Professeur Michael Fayon, le Professeur Pascal Barat, le Docteur Pascal Pillet, le Docteur Claude Billeaud et Mme Fabienne Nacka, dont l'engagement et la passion dans le service de santé public ont largement contribué à soutenir cette démarche de recherche.

entre les patients et leur médecin ?

H1. Le paysage informationnel en matière de santé a subi une dérégulation substantielle en raison de la démocratisation d'Internet. Désormais, les professionnels de la santé ne sont plus les seuls acteurs autorisés à délivrer de l'information médicale. Ils sont confrontés à la concurrence de nouveaux entrants, souvent plus agiles dans leur utilisation des stratégies de communication en ligne. Cette évolution nécessite une réévaluation de la place et de l'influence des professionnels de la santé dans le monde numérique.

H2. La logique du marketing digital, notamment le référencement sur les moteurs de recherche, a modifié la dynamique de la notoriété et de l'influence dans le secteur de l'information de santé. Ces transformations ont souvent pour conséquence de niveler la qualité des informations disponibles, en privilégiant l'effet sensationnel. Les stratégies informationnelles et communicationnelles doivent donc être analysées à la lumière de ces nouveaux mécanismes de visibilité et d'influence.

Dans ce contexte, l'étude de la sociologie des acteurs numériques de l'information de santé est cruciale non seulement du point de vue de la recherche en sciences sociales, mais aussi du point de vue pratique. Comprendre ces dynamiques peut aider à identifier les risques et les opportunités que présente la communication numérique pour les acteurs de la santé en France.

## **Plan du rapport**

Le but du projet de recherche est de saisir la diversité et la complexité des discours et des stratégies discursives adoptées par les patients et leurs accompagnants lorsqu'ils consultent des informations de santé en ligne. L'objectif est de cartographier l'écosystème numérique, d'identifier les acteurs, de comprendre leur positionnement, d'évaluer leur influence et de décrypter leur rôle dans la prescription d'informations de santé. Plus précisément, le protocole d'étude vise à démontrer la faisabilité d'une analyse globale de cet écosystème en se concentrant sur les maladies chroniques chez l'enfant comme étude de cas. La méthodologie de ce projet de recherche se déroulera en trois phases distinctes :

1. Une revue théorique initiale permettra de situer le projet dans le contexte des recherches et des théories qui ont abordé l'information de santé en contexte numérique.
2. Une étude méthodologique suivra pour expliciter les choix épistémologiques et techniques qui sous-tendent le fonctionnement du dispositif d'analyse.
3. Enfin, une étude comparative sur deux cas, celui du diabète chez l'enfant et celui de l'asthme pédiatrique, fournira une illustration pratique de l'application des méthodes et des outils du projet.

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

L'avènement du web a radicalement transformé le paysage de l'information en matière de santé, modifiant profondément la relation entre les patients, les professionnels de la santé et les institutions médicales. Dans cette nouvelle dynamique, Internet ne sert plus uniquement de dépôt d'informations, mais se transforme en une véritable plateforme d'interaction et de co-construction du savoir. La présence des patients dans ces espaces numériques a des conséquences complexes sur le dialogue médical et la gestion des maladies. Ainsi, le web est devenu un territoire d'étude où se cristallisent des enjeux majeurs en sciences de l'information et de la communication (SIC) et en santé publique.

La problématique centrale de cette recherche vise à examiner comment le web réorganise les relations et les discours autour de l'information de santé. L'étude cherche à comprendre l'hybridité discursive qui caractérise ces espaces, combinant des dimensions scientifiques, personnelles et affectives. Dans cette première partie, nous avons cherché à confronter les méthodologies et les perspectives épistémologiques de la santé publique et des Sciences de l'Information et de la Communication (SIC) afin d'établir un dialogue interdisciplinaire. La question fondamentale est donc : comment peut-on élaborer une méthodologie interdisciplinaire pour étudier de manière exhaustive et nuancée l'impact du web 2.0 sur l'information de santé ?

La recherche se structure autour de trois axes principaux :

1. L'information de santé en santé publique : Dans cette section, nous traiterons des défis méthodologiques rencontrés en santé publique et en SIC lors de l'étude de l'information de santé sur Internet. Nous identifierons les limites et les potentiels de chaque approche, tout en mettant en évidence leur complémentarité.
2. La Patientsphère en ligne : vers un individualisme communautaire : Ce segment explorera le rôle et l'impact des patients dans ces espaces numériques. Nous examinerons en particulier le rôle des associations de patients et les formes d'engagement qui émergent, allant d'un individualisme collectif autorégulé à des formes plus structurées d'action collective.
3. L'apport de la sociologie du web à la question de l'information de santé sur Internet : Dans cette dernière section, nous nous concentrerons sur ce que les études en sociologie du web peuvent contribuer à l'analyse des discours et des pratiques autour de l'information de santé. Cette section propose des pistes de réflexion sur les nouvelles formes de régulation, de co-construction du savoir et de négociation de l'autorité.



Dans cette section, notre objectif est de résumer les enseignements clés issus de cette recherche, tout en envisageant des perspectives interdisciplinaires futures pour approfondir notre compréhension de l'impact du web sur l'information de santé. En adoptant cette approche, nous aspirons à créer un lien entre les perspectives méthodologiques de la santé publique et des Sciences de l'Information et de la Communication (SIC), qui semblent divergentes, afin de contribuer à une meilleure compréhension des transformations engendrées par le web 2.0 dans le domaine de la santé.

## **I.1 - Recherches en santé publique, l'évaluation de la qualité de l'information à l'aune des pratiques médicales**

La question de l'évaluation de la qualité de l'information médicale en ligne se situe dans un cadre épistémologique principalement défini par les chercheurs en santé publique, affiliés à la section 46 de la nomenclature académique. Ces recherches sont généralement conçues comme des efforts herméneutiques visant à mesurer la conformité des corpus numériques aux normes préétablies de la publication médicale, en utilisant des critères d'évaluation formels au sein d'une échelle de qualité définie. L'approche méthodologique adoptée dans ces études se caractérise par une rigueur quantitative formalisée qui présente une stabilité transversale remarquable à travers les articles. Cette uniformité méthodologique est non seulement le résultat, mais aussi le produit d'une homogénéité disciplinaire. Cette dernière prédispose à une réplique constante des paradigmes d'analyse appliqués à des corpus hétérogènes. Autrement dit, l'isomorphisme disciplinaire induit une forme de dépendance au chemin qui tend à propulser les mêmes modèles d'analyse à l'interface de différents domaines de pathologie. L'écho méthodologique récurrent dans ces études, observé au cours des quinze dernières années, a produit des conclusions d'une cohérence singulière malgré la multitude de pathologies examinées. En effet, ce domaine de recherche a généré des perspectives cohérentes sur des pathologies aussi diverses que les tumeurs cérébrales pédiatriques (Hargrave et al., 2003), la neuro-oncologie (Lévêque et al., 2007), l'oncologie otorhinolaryngologique (Evrard et al., 2009), la péridurale en obstétrique (Espitalier et al., 2013 et 2014), la diversification alimentaire (Banti et al., 2016), la question du gluten (Banti et al., 2017), ou encore le dépistage des cancers oraux (Vivien et al., 2017), entre autres. Cette concordance des résultats à travers les pathologies, par ailleurs corroborée par des variantes américaines de ce modèle de recherche, suggère une transposabilité méthodologique qui, tout en renforçant la validité interne des études, soulève inévitablement la question de leur capacité à s'adapter aux changements épistémologiques et aux déplacements paradigmatiques qui traversent actuellement la discipline de la santé publique. En résumé, l'évaluation de la qualité de l'information de santé sur Internet est une catégorie de recherche qui, bien que solide dans ses conclusions, doit faire face aux défis inhérents à toute tentative de normalisation épistémique dans un domaine en constante évolution, marqué par les dynamiques complexes des technologies de l'information et de la communication, ainsi que par les changements sociétaux qui influencent la perception et l'utilisation de l'information

médicale en ligne.

D'un point de vue méthodologique, le protocole quantitatif utilisé par ces études démontre une rigueur indéniable, mais aussi une portée conceptuelle limitée. La mise en œuvre de la recherche se fait par l'introduction d'une série de termes spécifiques tirés du lexique médical spécialisé dans deux ou trois moteurs de recherche algorithmiques. Ce choix méthodologique révèle une omission épistémologique significative : l'ignorance des moteurs de recherche en tant que dispositifs médiatiques et technologiques influençant la présentation de l'information, notamment par le biais de l'adresse IP, de l'historique de navigation et de la géolocalisation de l'utilisateur. Ces écarts méthodologiques entraînent des biais dans la constitution du corpus, limitant ainsi la portée des protocoles de recherche. De plus, la construction même de ce corpus prédispose à une surreprésentation de la connaissance médicale spécialisée, en raison de la technicité des termes-clés utilisés. Ainsi, l'élimination substantielle — souvent dépassant les 85 % — des sites jugés «non pertinents» ou «sans intérêt» effectuée, le plus souvent par des médecins spécialisés responsables de l'évaluation, soulève des questions sur la population des énoncés étudiés. Ce que l'on observe alors est la création d'un corpus de nature principalement médicale, sélectionné pour examiner la discoursivité médicale sur le web et pour l'évaluer en fonction des normes régissant le champ médical, telles que les sociétés savantes et les recommandations cliniques

En agissant ainsi, des espaces discursifs alternatifs, tels que les forums et blogs de patients, sont exclus du corpus, leurs contributions étant jugées périphériques voire sans intérêt, bien qu'elles occupent un espace significatif sur le web. Pour l'évaluation de ces corpus, les chercheurs s'appuient sur des instruments d'évaluation épistémologiquement ancrés et institutionnellement validés qui mesurent la confiance dans l'information médicale (les grilles DISCERN, HON, NetScoring, ou encore les normes établies par la Commission Européenne). Il n'est pas rare que ces chercheurs synthétisent ces différents référentiels pour élaborer leurs propres grilles d'évaluation, une démarche illustrée entre autres par Frémont et al. (2001) et Espitalier et al. (2013). Les sites ou pages web ainsi isolés sont examinés selon des critères éditoriaux précis, visant à mesurer leur transparence et la sincérité de leurs objectifs. Ceux-ci incluent la citation des références, la date de publication et de mise à jour, l'identification des auteurs, ainsi que la clarté des objectifs pragmatiques et commerciaux. De plus, une évaluation de la qualité du contenu est entreprise, en considérant des variables telles que la pertinence de l'information, son accessibilité, la gestion des incertitudes liées à la maladie ou au traitement, les modalités d'intervention, les effets secondaires, les alternatives thérapeutiques et d'autres dimensions impactant la vie quotidienne du patient. À la lumière de ces choix méthodologiques, il est évident que la question de recherche sous-jacente, ou plutôt le biais qui informe ces méthodologies, réside dans l'évaluation de la qualité de l'information provenant des acteurs médicaux eux-mêmes. Ce qui en découle est une sorte d'"endogamie épistémologique" qui semble régir les actions des médecins-chercheurs en santé publique, une réification du savoir médical qui pourrait menacer

l'écologie informationnelle de la santé sur le web. La question que posent les chercheurs en santé publique se résume au rôle des médecins eux-mêmes sur Internet.

Les conclusions tirées de ces études semblent souvent fortement influencées par les choix méthodologiques effectués. Il n'est pas surprenant que les corpus non médicaux soient disqualifiés en tant que "bons" corpus médicaux lorsque les critères d'évaluation sont définis à l'avance selon les normes médicales. Ce faisant, l'hégémonie discursive du corps médical est non seulement reproduite, mais aussi naturalisée, occupant souvent plus de la moitié du corpus analysé. Les associations de patients et les patients individuels sont relégués à la périphérie, ne représentant que près de 10% du corpus dans la plupart des études. L'écosystème discursif qui émerge est pour le moins schizophrénique : d'un côté, des professionnels de santé s'adressant principalement à d'autres professionnels de santé ; de l'autre, un ensemble hétérogène d'acteurs s'adressant directement aux patients. Malgré cette diversité apparente, il y a un consensus parmi les chercheurs sur l'insuffisance de la qualité de l'information de santé disponible : à peine un tiers des pages analysées sont considérées comme satisfaisantes, tandis que les deux tiers restants sont jugés médiocres, voire carrément inadéquats. Bien que les informations dangereuses soient rares, la qualité globale du corpus est généralement jugée médiocre. Cette critique est renforcée par des études longitudinales qui soulignent la volatilité des corpus d'année en année, avec plus de la moitié des sources initialement identifiées disparaissant des moteurs de recherche pour des requêtes identiques. Lorsque la parole du patient n'est pas explicitement disqualifiée comme dénuée d'intérêt - une attitude manifeste dans des travaux comme ceux d'Espitalier (2013) -, elle est analysée principalement à travers le prisme du discours médical. Le patient est conceptualisé comme une entité passive, en quête d'informations médicales standardisées concernant la nature de sa maladie, les options de traitement et l'interprétation des diagnostics médicaux, une perspective déjà soulignée par Akrich et al. (2009).

Le panorama esquissé par ces travaux s'avère singulièrement riche et complexe, mettant en lumière des enjeux qui dépassent largement les préoccupations médicales traditionnelles. La reconnaissance, même partielle, de l'émergence d'acteurs tiers dans le domaine de l'information en santé en ligne par les professionnels médicaux eux-mêmes, constitue un tournant épistémologique. Cette prise de conscience des contributions discursives des patients via les forums et blogs, ainsi que des associations de patients, introduit une déstabilisation salutaire dans un champ autrement dominé par une homogénéité médicale. Bien qu'un seul article (Legros, 2010) en santé publique valorise explicitement cette diversification des voix via le web 2.0, la reconnaissance des compétences communicationnelles de ces nouveaux acteurs par les médecins traduit un assouplissement du monopole épistémique traditionnellement exercé par le monde médical et une acceptation croissante de la parole profane du patient et/ou de son entourage. Un autre élément à souligner est la remise en question implicite des hiérarchies informationnelles telles qu'établies par les moteurs de recherche. La qualité de l'information médicale se révèle, dans ces études, comme étant complètement décorrélée du

rang dans les moteurs de recherche ou du nombre de citations. Ce constat met en lumière la problématique plus large de la mécanique algorithmique et de l'économie du marketing web, qui régulent la visibilité et donc l'impact potentiel des informations en santé en ligne. Néanmoins, il est à noter que ces travaux, tout en éclairant certaines zones d'ombre, révèlent aussi leurs propres limitations. Ils illustrent l'importance d'un questionnement disciplinaire, mais également l'impératif d'une exploration pluridisciplinaire pour pallier les inévitabilités du champ d'invisibilité qui résulte des partis pris méthodologiques et épistémologiques. En d'autres termes, ce que ces recherches mettent en exergue, c'est à la fois la nécessité et l'urgence d'une remise en question des paradigmes dominants. Cela passe par la prise en compte de la pluralité des voix et des savoirs, sans laquelle la construction d'un corpus véritablement inclusif et représentatif demeure une chimère. L'ensemble de ces observations suggère une tension inhérente entre l'ouverture aux discours marginaux et la préservation d'un certain niveau d'expertise médicale. Cette tension, loin d'être résolue, constitue un enjeu majeur dans le débat sur la démocratisation de l'information en santé et sur le rôle des différents acteurs impliqués. Elle soulève des questions d'éthique de la connaissance, de justice épistémique et de légitimité discursive, questions qui exigent un élargissement du prisme analytique pour englober des considérations tant sociologiques que technologiques, économiques et éthiques. En somme, le défi est celui de la construction d'un nouveau modèle épistémologique capable de réconcilier les exigences contradictoires d'un environnement informationnel en constante évolution. En somme, ce que l'on observe est une forme de réductionnisme médicaliste qui non seulement encadre mais aussi restreint le spectre de la parole légitime dans l'univers de l'information de santé sur internet. Ce phénomène reflète une forme d'endogamie épistémologique qui renforce une écologie informationnelle asymétrique, marginalisant les acteurs non médicaux et leurs contributions discursives au débat sur la santé. Ainsi, l'état actuel des recherches en santé publique invite à une refonte méthodologique qui prendrait en compte la pluralité des voix et des savoirs dans l'élaboration de corpus plus inclusifs et représentatifs. Nous voulons souligner ici les conclusions annexes qui nous semblent riches d'enseignements dans ces études. Les professionnels de santé reconnaissent une montée en puissance des acteurs tiers sur internet. Ils soulignent la forte présence de la parole du patient à travers le web 2.0 (forums et blogs) et des associations de patients qui les représentent. Mais ils soulignent en même temps leur incapacité à les étudier du fait même de leur cadrage disciplinaire. Nous n'avons trouvé qu'un article en Santé publique où le web 2.0 a grâce devant le chercheur (Legros, 2010). Pourtant dans l'ensemble, ils reconnaissent à ces nouveaux acteurs les mêmes qualités de communication que les institutions de santé (université, CHU ou institution de santé). Ils constatent aussi la montée de la presse grand public qu'elle soit généraliste ou spécialisée en santé. Ils montrent par ailleurs que la qualité de l'information médicale est complètement décorrélée du rang dans les moteurs de recherche ou du nombre de citations. Ils touchent alors du doigt la problématique de l'économie marketing web qui structure la production et la distribution de l'information en ligne. On voit alors comment l'ensemble de ces travaux montre l'apport et l'intérêt d'une approche disciplinaire qu'il faut saluer. Mais aussi comment

les partis pris produisent un champ d'invisibilité qui ne peut être dépassé que par une exploration pluridisciplinaire.

## **I. 2 - Sociologie de la communication : A la recherche des nouveaux régimes discursifs du patient en ligne.**

Dans cette seconde partie, nous allons mettre en dialectique un autre aspect de la recherche en information de santé : les nouveaux régimes discursifs du patient en ligne. Nous adoptons une perspective qui diffère de manière significative des paradigmes de la discipline 46 sur l'analyse de contenu médical. Ici, l'accent est mis résolument sur les espaces numériques web 2.0 — principalement les forums et les blogs — où la parole du patient est non seulement amplifiée, mais aussi substantiellement modifiée dans ses modalités discursives. Dans cette optique, l'approche qualitative surpasse l'approche quantitative, offrant une analyse du discours qui va au-delà de la simple métrique pour embrasser une herméneutique sensible aux subtilités du langage et des interactions. Cette orientation repose sur l'hypothèse que l'émergence de nouveaux acteurs discursifs, principalement grâce à la démocratisation de l'accès à Internet, a entraîné une modification substantielle dans la production du savoir sur la santé et, par extension, sur la dynamique de la relation patient-soignant. Cet axe de recherche, non médicalisé et clairement ancré dans les sciences humaines et sociales (SHS), provient principalement de deux centres intellectuels en France. D'une part, il y a le Centre de Sociologie de l'Innovation (CSI) de Mines ParisTech, dont les fondements ont été posés par Michel Callon et qui est actuellement dirigé par Madeleine Akrich. Les travaux issus de ce centre s'inspirent largement de la sociologie pragmatique et des théories de la traduction de Callon, utilisant l'information de santé comme un champ expérimental pour leurs modèles théoriques (Akrich et Méadel, 2002 et 2007; Callon et Rabeharisoa, 2008; Akrich et al., 2009 et 2014; Méadel, 2009). D'autre part, on note une concentration significative de chercheurs en Sciences de l'Information et de la Communication (SIC), spécifiquement au sein du GRESEC (Groupe de recherche sur les enjeux de la communication de l'Université Grenoble Alpes). Ces derniers contribuent à l'autre moitié de la littérature scientifique sur cette question en France et incluent des chercheurs tels que Convert et Demailly (2003), Lafon et Paillart (2007), Romeyer (2008, 2010, 2012), Staii et al. (2010), et Paganelli et al. (2014), parmi d'autres. Ce réalignement disciplinaire autour de la sociologie de la communication esquisse les contours d'un nouveau régime discursif du patient en ligne, caractérisé par une porosité accrue entre savoirs experts et savoirs profanes, entre discours institutionnels et discours issus du vécu. Cela indique une évolution notable dans les pratiques et les paradigmes qui régissent l'information de santé en ligne, tout en ouvrant un vaste champ d'investigation pluridisciplinaire.

La conjoncture complexe dans laquelle s'inscrivent ces études invite à un examen critique des paradigmes méthodologiques et épistémologiques. Ceux-ci sont souvent fondés sur un interactionnisme pragmatique, qui, bien que pertinent pour une étude qualitative des échanges en

ligne, rencontre des obstacles inhérents à la spécificité du milieu numérique. Premièrement, l'inadéquation entre la constitution du corpus et les exigences d'exhaustivité et de représentativité laisse une marge d'imprécision substantielle. Le flou méthodologique est souligné par l'utilisation de termes tels que "études exploratoires" et "pré-recherche", qui trahissent une certaine réticence à universaliser les résultats en raison de l'impossibilité de définir une population de référence ou un échantillon représentatif. Cela suggère une lacune dans la méthodologie de la recherche numérique, révélant un besoin urgent d'outils méthodologiques plus sophistiqués et adaptés (Cormerais et Lakel, 2018; Proulx, 2018). Deuxièmement, l'extraction des données elle-même est problématique. L'approche peut varier, de l'extraction de posts sur une période donnée à la comparaison de blocs de discussions dans des listes de diffusion, chacune avec ses avantages et ses inconvénients, y compris des questions éthiques relatives à l'observation invisible (Bettaïa, 2016). Troisièmement, la diversité intrinsèque des forums et des listes de diffusion complique encore plus l'analyse. L'architecture communicationnelle varie sensiblement d'une plateforme à une autre, remettant en question l'extensibilité des découvertes d'une étude à l'autre. Plusieurs chercheurs postulent une différenciation fondamentale entre les listes de discussion, considérées comme des milieux informels de communautés aux rôles distribués, et les forums, vus comme des espaces plus transitoires de questions-réponses. Toutefois, des recherches plus récentes suggèrent que cette distinction peut être plus nuancée, surtout quand elles s'appuient sur des monographies de forums spécifiques. En somme, les approches actuelles en sociologie de la communication se heurtent à des défis méthodologiques majeurs qui entravent la progression vers une compréhension exhaustive des nouveaux régimes discursifs en ligne. Elles oscillent entre le descriptif et l'exploratoire, ménageant ainsi un espace de recherche qui, tout en étant extrêmement riche en perspectives, est également caractérisé par son instabilité épistémologique. La mise en relief des séquences d'interactions à travers le prisme goffmanien et l'analyse lexicale du discours contribuent à une cartographie préliminaire mais insuffisante des dynamiques en jeu. Il devient donc impératif de développer de nouvelles stratégies de recherche qui réconcilient les approches qualitatives avec les exigences d'une science de la communication de plus en plus inscrite dans des dynamiques multidisciplinaires et transversales.

La cristallisation de ces recherches autour de trois axes stratégiques — le plan stratégique, le plan pragmatique et le plan stylistique — signifie que la relation patient-soignant en ligne ne peut être décrite par une approche unidimensionnelle. Elle opère plutôt dans un maillage complexe de variables qui s'articulent de diverses manières. Sur le plan stratégique, l'accent est mis sur l'agentivité du patient, manifestée à travers la prise de parole en ligne. Ce discours s'inscrit dans un espace performatif où la parole est non seulement une expression de l'expérience individuelle, mais aussi un outil d'empowerment. Les espaces numériques deviennent alors des lieux de parole démystifiés, moins hiérarchisés et moins anxiogènes que la relation médicale traditionnelle. La notion d'empowerment interpelle ainsi directement les questions de pouvoir, de hiérarchies et de distributions de rôles dans la

structure institutionnelle médicale. Cette ligne d'interrogation renvoie aux mouvements associatifs des patients dans les années 1980 en France et invite à penser la manière dont les dynamiques de pouvoir sont redistribuées dans le cadre numérique. Le plan pragmatique interroge la construction collective du parcours de soins. Dans des communautés numériques de patients, l'agrégation des expériences et des savoirs individuels pourrait alimenter des projets collectifs, redéfinissant ainsi le parcours de soins comme une entreprise co-construite. Cette dynamique résonne avec les travaux de Parsons et Strauss sur la relation soignant-soigné, en remplaçant leur schéma hiérarchique par un modèle co-constructif numérisé (Castel, 2005; Paganelli et al., 2014). Le plan stylistique, enfin, envisage une formalisation progressive d'un savoir expérientiel propre à la communauté de patients en ligne. Ce savoir servirait alors de fondement légitimant le "patient expert," capable de co-construire son propre parcours de soins. Mais, bien entendu, ces dynamiques ne sont pas sans implications sur la gouvernance de ces espaces numériques. La liberté apparente de ces forums et listes de discussion est souvent soumise à une régulation informelle, voire soft, orchestrée par ce que l'on peut appeler des « leaders discursifs » (Lakel et Massit Folléa, 2007). Ceux-ci utilisent des stratégies de "coersédution" pour maintenir un certain ordre discursif, pouvant aller de l'ignorance manifeste à des mesures plus drastiques telles que le bannissement (Vercher et Touboul, 2010). Ce complexe enchevêtrement d'axes d'investigation s'inscrit clairement dans une dynamique de recherche pluridisciplinaires. Il exhorte à une réflexivité accrue sur la méthodologie et l'épistémologie dans le domaine de la sociologie de la communication en ligne. Il invite également à explorer les possibles translations et transformations des formes de socialisation des patients qui existent hors-ligne. À cet égard, une analyse comparative entre les modes de socialisation en ligne et hors-ligne peut offrir un éclairage précieux sur les nouvelles formes d'échanges et de rôles émergents, enrichissant ainsi le corpus scientifique de ce domaine en constante évolution.

### **I.3 - Ce qu'apporte la sociologie du web à la question de l'information de santé sur internet.**

La sociologie du Web, conjointement aux Sciences de l'Information et de la Communication (SIC), apporte un regard novateur sur les dynamiques informant la relation patient-soignant dans le contexte d'une information de santé de plus en plus digitalisée. Le rôle des associations de patients comme instances médiatrices dans le système de santé est particulièrement instructif. Ces entités, souvent investies d'une double légitimité — représentative des intérêts des patients et médiatrice de la politique de santé publique —, effectuent un travail complexe de traduction, d'évaluation et de participation dans l'espace de la communication en santé. Ces associations agissent en quelque sorte comme des nœuds dans un réseau d'information où les patients ne sont plus de simples « prescrits », mais deviennent des acteurs à part entière dans la coconstruction du savoir médical. Ce basculement vers un modèle participatif ne transforme pas les espaces numériques en lieux de mobilisation

formelle, mais en espaces ouverts pour un "individualisme collectif autorégulé" (Convert et Demailly, 2003). Dans cette architecture, chaque individu est souverain dans son parcours d'information et de soins, tout en étant influencé par une dynamique collective d'échange d'informations et de meilleures pratiques. L'importance de ces « forums hybrides ouverts » (Barthes, Callon et Lascoumes, 2014) réside en partie dans leur capacité à moduler la critique de la pratique médicale. Si l'institution médicale en tant que telle est rarement remise en cause — souvent défendue, surtout lorsqu'elle est mise en parallèle avec des alternatives non conventionnelles —, l'exercice individuel d'un médecin en particulier peut être soumis à un examen plus minutieux. Cela révèle une distinction nuancée, mais cruciale : le consensus autour de la science médicale ne se traduit pas nécessairement par une confiance aveugle envers tous les praticiens. Au contraire, la démocratisation des espaces d'information permet un contrôle social de la compétence médicale qui était auparavant souvent confiné aux cercles professionnels et institutionnels. Ces nouvelles formes d'engagement, de critique et de contrôle social se déroulent dans un espace numérique qui est fondamentalement un lieu public non contraignant. Il se situe dans un espace intermédiaire, entre des collectifs formellement structurés et des actions individuelles. Cette position liminale invite à repenser les termes de l'engagement des patients dans les politiques de santé, qu'elles soient publiques ou privées. La nature hybride de ces interactions — entre individualisme et collectivité, entre critique et conformité, entre liberté et responsabilité — appelle des méthodologies de recherche et des cadres d'analyse qui sont à la fois nuancés et adaptables. En somme, la sociologie du Web et les SIC nous incitent à une compréhension plus intégrée et complexe de la façon dont les technologies de l'information transforment la santé et les soins. Ces évolutions ne sont pas simplement des changements d'échelle ou de rapidité dans la diffusion de l'information, mais des transformations qualitatives qui influencent les dynamiques du pouvoir, les formes de légitimité, et même les paradigmes épistémologiques qui sous-tendent le système de santé.

Les espaces numériques émergent comme des scènes de corégulation informationnelle plutôt que comme des zones de contestation radicale du pouvoir médical. Ils agissent comme des interfaces entre les mondes médicaux institutionnalisés et les patients, servant de terrains de délibération sur des thématiques médicales diverses. Contrairement à la dissidence, ils incarnent une forme d'adjonction dialogique au système de santé établi, offrant des espaces de débat et d'élaboration de savoirs alternatifs ou complémentaires. Il est crucial de noter ici que la médiation opérée par ces espaces n'est pas un simple acte de traduction des connaissances médicales en un langage profane. Elle constitue également un effort d'évaluation et d'expérimentation des savoirs profanes, notamment ceux qui sont aux marges du biologisme médical dominant. Ce rôle de médiation s'apparente ainsi à celui que les associations de patients ont traditionnellement joué comme caisses de résonance du diagnostic médical, mais sans la sacralité institutionnelle qui les accompagne. Des alternatives thérapeutiques telles que l'homéopathie, la méditation, l'hypnose, ou encore la phytothérapie et les régimes



spécifiques, viennent compléter le spectre des options de soins en discussion. S'ils ne se positionnent pas comme des remplacements à la médecine conventionnelle, ces traitements relèvent de logiques de soin qui échappent souvent aux paramètres strictement biologiques et cliniques. Leur présence dans ces forums atteste d'une curiosité collective et d'un esprit d'expérimentation qui est à l'initiative des patients eux-mêmes. Ce phénomène révèle une subversion latente du modèle médical en vigueur, non pas par son rejet, mais par son extension et sa complexification<sup>2</sup>. Il est également digne de note que l'émergence de ces thématiques dans les délibérations de sociétés savantes et d'institutions médicales témoigne de l'influence croissante de ces espaces numériques sur la pratique médicale. Ceci pourrait signaler une nouvelle forme de « science citoyenne » où le patient n'est plus seulement le sujet passif d'un diagnostic ou d'un traitement mais devient un cocréateur, voire un cocreateur du système médical en vigueur. Dans ce sens, on assiste à une recodification des frontières entre savoir expert et savoir profane. Ce phénomène peut être conceptualisé comme une forme de « polyphonie épistémique », où différents régimes de vérité coexistent et interagissent de manière plus ou moins harmonieuse. L'existence même de ces forums est un symptôme de l'insuffisance du système de soins actuel à répondre à la complexité et à la multidimensionnalité des besoins et des aspirations des patients, et suggère la nécessité d'une révision ontologique et épistémologique des pratiques médicales. L'équilibre entre expertise médicale et expérience patient doit être constamment négocié et reconstruit. En somme, les espaces numériques de corégulation informationnelle, loin d'être de simples canaux de diffusion, deviennent des laboratoires sociaux où se négocient les futurs de la médecine, et par extension, les paradigmes. L'émergence d'un nouveau régime discursif dans les espaces numériques consacrés à la santé constitue un terrain d'investigation fertile pour les sciences humaines et sociales (SHS). Le recours massif au témoignage personnel et au partage d'expériences vécues individuellement, tout en soulignant la singularité de chaque cas, inscrit ces espaces dans une logique d'hybridité informative à la croisée du personnel et du scientifique. Il s'agit là d'un régime du "pathos" où les données émotionnelles cohabitent, et même se complètent, avec des informations de nature plus objective. Cet échange d'expériences s'appuie sur une forme discursive spécifique, marquée par une certaine désinhibition linguistique, qui rend manifeste la primauté de l'expérience vécue sur les conventions langagières et textuelles. Cette hybridité discursive introduit un bouleversement épistémologique et une diversification des régimes de vérité, car elle s'articule autour d'une "économie du témoignage" (Battaïa, 2013; Paganelli et al., 2014). Cette économie présuppose que la reconnaissance de la singularité des cas individuels n'est pas antinomique à une visée plus générale de compréhension et de traitement des maladies. Il s'agit en effet d'une démocratisation du discours sur la santé où le particulier est subsumé dans le collectif non pas par effacement, mais par amplification de sa singularité. L'anonymat qui préside souvent à ces témoignages introduit une dimension d'égalité et de

---

<sup>2</sup> Il suffit de penser à la place que des pratiques jugées il y a peu comme à la limite du charlatanisme sont devenu aujourd'hui des outils légitimes de la pratique médicale : médecine chinoise, herbothérapie, hypnose, gestions du stress et de la douleur, etc.

liberté de parole, tout en maintenant un certain ordre social sous-jacent. Cette dynamique témoigne de la présence d'une "nétiquette" souvent implicite, qui organise les rapports sociaux et hiérarchiques au sein de ces communautés. On assiste ainsi à une forme de régulation communautaire qui s'opère selon des axes multiples: le degré d'expertise, le niveau d'engagement dans la communauté, et la qualité de la contribution au savoir collectif (Paganelli et al., 2014). Ce régime de régulation est complexe, car il conjugue à la fois des règles tacites et explicites, établissant une forme d'auto-régulation. Il est marqué par des formes hiérarchiques fluides, allant du "lurker" silencieux à l'expert reconnu, en passant par l'amateur et l'engagé. Il en ressort que ces espaces numériques deviennent des terrains de constitution de nouvelles formes de "biopolitique communautaire" où la gouvernance de la vie et de la santé s'élabore non plus uniquement à partir des institutions médicales, mais aussi à partir des communautés elles-mêmes. Ces régimes discursifs et hiérarchiques internes sont donc des lieux d'exercice de nouvelles formes de pouvoir et de résistance, redéfinissant les dynamiques de l'autorité médicale, de la communauté de patients, et de l'État. En somme, ce nouveau régime discursif matérialise une épistémologie plurielle de la santé, en intégrant des dimensions jusqu'alors reléguées au second plan. Il repose sur une logique participative et émancipatrice qui renégocie les frontières traditionnelles du savoir médical, et fait place à une multiplicité de voix et d'expériences. Ce faisant, il éclaire les impensés et les angles morts du système médical traditionnel, et offre des pistes pour une médecine plus inclusive, réflexive et démocratique.

Dans cette partie, nous avons examiné la complexité du Web 2.0 en tant qu'espace de dialogue entre patients, praticiens et institutions médicales. Nous avons mis en lumière l'hybridité discursive caractérisant ces espaces, marqués par des régimes de vérité multiples et l'émergence de nouvelles formes de biopolitique communautaire. À partir des théories en sciences humaines et sociales, nous avons observé comment le discours qui y circule révèle à la fois des dimensions scientifiques, personnelles et affectives. Ces espaces numériques se transforment en chambres de discussion communautaires, mais également en plateformes d'expression de soi, de sa souffrance et de sa vie. Ils redéfinissent aussi la dynamique du pouvoir et de l'autorité, notamment médicale, en intégrant divers degrés d'expertise et d'engagement communautaire. Tenant compte des constats de la santé publique, qui déplore la piètre qualité du corpus d'informations de santé sur Internet, et des sciences de l'information et de la communication, qui y voient au contraire un terrain de recherche fascinant, il est essentiel de repenser la méthodologie pour aborder ces espaces numériques. En fait, les deux disciplines, bien qu'en opposition méthodologique, se complètent dans leurs constats. La santé publique y voit une faiblesse structurelle, tandis que les sciences de l'information et de la communication y voient une richesse pragmatique et discursive. Cette dichotomie souligne la nécessité de travaux interdisciplinaires qui sauront intégrer les forces et les faiblesses de chaque approche. En termes de perspectives, la première est celle de la constitution d'un corpus de recherche qui serait à même de répondre aux exigences de rigueur de la santé publique tout en respectant la complexité

inhérente à l'espace du Web 2.0. Cela nécessiterait un réexamen des méthodologies classiques de collecte de données et de leur analyse afin de proposer des approches plus nuancées et polyvalentes. La deuxième perspective est celle de l'intégration de ces observations dans les politiques de santé et les pratiques médicales. Accepter une conception plus large de l'information de santé, qui inclut l'expérience personnelle du patient et les savoirs communautaires, pourrait offrir des voies pour améliorer la performance des parcours de santé. En définitive, ces espaces numériques ne sont pas simplement des forums d'échange d'informations médicales, mais des lieux de redéfinition du rôle du patient, du médecin et de l'institution médicale dans la co-construction du savoir et de l'expérience de la maladie. Une approche interdisciplinaire, enrichie par des contributions à la fois de la santé publique et des sciences de l'information et de la communication, s'impose pour saisir pleinement cette complexité.

## II – Un dispositif d'intelligence numérique : enjeux méthodologiques

L'importance du Web en tant que source d'information essentielle n'est plus à démontrer. Devenu, au fil des années, l'espace de publication préféré des organisations et des individus, Internet est aujourd'hui une bibliothèque de publications numériques qui a depuis longtemps fait exploser les seuils de l'archivage, tels que nous les concevions avant la numérisation de la communication. Cependant, loin d'être une aubaine pour les chercheurs et les experts, le Web, en tant qu'espace ouvert, est une source d'information plus que problématique. Les enjeux liés au traitement des données, l'hétérogénéité des documents et des situations communicationnelles, ainsi que la difficulté d'identifier les sources rendent le Web similaire à une gigantesque boîte aux lettres remplie à ras bord de prospectus, d'ouvrages savants et de catalogues de vente par correspondance, sans aucune organisation ni gestion des connaissances. Néanmoins, cet abîme où s'entassent toutes les prises de parole du monde nous confronte à un phénomène unique dans l'histoire de l'humanité : la possibilité, du moins pour l'instant théorique, pour tout un chacun d'accéder à l'immense production des discours publics produits en n'importe quel point du globe.

Le Web est donc un paradoxe. En tant que gigantesque archive, il demeure toutefois difficilement utilisable comme source de documentation. Face à cette infobésité, les acteurs sociaux, et en premier lieu les professionnels de l'information, ont dû rapidement s'organiser pour développer des outils d'intermédiation entre l'archive numérique et les besoins des lecteurs. L'économie numérique a été, avant tout, une économie d'accès à l'information. D'abord, les moteurs de recherche, dont le marché semble aujourd'hui stabilisé autour du monopole de Google, n'ont eu de cesse de rechercher les documents (à l'aide de crawlers) et de les classer (via le ranking) sur la base d'algorithmes sémantiques dont la complexité, protégée par des brevets, constitue toute la valeur de ces grands groupes de la nouvelle économie. Sur la base d'une requête composée de mots-clés, l'enjeu était de classer les documents les plus pertinents au regard d'un questionnement. Ensuite, ce fut au tour des "social bookmarkers" de nous fournir des plateformes de gestion — en réalité, de mise en mémoire — de nos favoris, de partage de ces favoris (on parle ici de curation) et de collaboration collective, permettant de s'appuyer sur l'intelligence de la foule pour classer le Web selon le succès de chaque document. À l'algorithmique sémantique, on ajoute une part de crowdsourcing et de mesure d'audience des documents pour définir, au-delà de la pertinence logique, une hiérarchisation de l'information en fonction de son usage social. Puis vint, dans un troisième temps, la révolution des réseaux sociaux — véritables "killer applications" de l'information — qui, non contents de réduire la granularité de la

publication à la conversation de bistrot, devinrent les outils de curation les plus puissants du Web, notamment Facebook et Twitter.

Tous ces infomédiaires du Web, que j'appellerai, dans le cadre du projet My Web Intelligence, "oracles" (car ce sont des instances auxquelles on soumet une requête et qui vous livrent une série de réponses), ont permis, dans un premier temps, d'indexer le Web selon des variables sémantiques, sociales et structurales (analyse des liens entrants et sortants). Cette indexation est au fondement de l'utilisabilité du Web par tout un chacun. Pour autant, les professionnels de l'information le savent bien : il faut un "travail" pour passer de la donnée à l'information, et de l'information à l'intelligence. Le design utilisateur de ces plateformes favorise avant tout cette fonction d'oracle delphique : je pose ma question, je reçois une réponse et je m'en vais. C'est dans cette limite et la nécessité de ce "travail" que se sont développées, ces cinq dernières années, des solutions de veille et d'intelligence du Web voire, plus spécialisées, des services de social media monitoring. L'intelligence numérique est le fruit d'un traitement et d'une gestion des données en ligne au service d'un travail de recherche.

## **II.1 – My Web Intelligence : enjeux des études numériques**

Le programme My Web Intelligence ambitionne de créer un dispositif intégré d'extraction (crawl), d'archivage, de qualification et de visualisation des données du Web, spécifiquement conçu pour catalyser les avancées dans le domaine des études digitales. Ce programme aspire à fournir aux chercheurs et aux experts, particulièrement ceux œuvrant dans le domaine de l'intelligence numérique et des humanités digitales, un outil permettant d'analyser finement les prises de parole et les interactions en ligne. La naissance et la croissance rapide de ce champ d'étude sont le reflet d'un contexte scientifique et économique extrêmement dynamique. Le data marketing et l'intelligence numérique ont en effet connu un développement exponentiel durant la dernière décennie. Initialement centrée sur le Web Analytics et le e-commerce, cette discipline a connu un élargissement substantiel de son spectre d'action avec l'apparition du Web 2.0 et notamment des réseaux sociaux. Ces derniers ont dû composer avec des enjeux de gestion de données massives pour développer un écosystème autour du Social Media Analytics, l'analyse ciblée des interactions sur ces plateformes. L'abondance des données générées a incité divers acteurs — agences de marketing digital, centres de recherche en intelligence économique et territoriale, et laboratoires universitaires — à développer des solutions d'analyse et de suivi des dynamiques conversationnelles sur les réseaux sociaux. Entre 2013 et 2015, l'investissement dans les start-ups proposant ces types de solutions a atteint plusieurs centaines de millions de dollars. L'importance de ces outils réside dans leur capacité à anticiper les attitudes et les préférences des consommateurs et à identifier proactivement les tendances émergentes relatives à une marque ou à un produit. Ils offrent ainsi un avantage stratégique et concurrentiel indéniable.

L'initiative du programme My Web Intelligence s'articule autour de la mise en place d'un outil analytique rigoureux destiné à l'étude des interactions en ligne. Cette démarche vise à enrichir les méthodologies existantes dans le champ des études numériques, en apportant un instrument capable de capturer la complexité et la nuance des discours et des échanges sur le Web. Le Web demeure un espace difficile à appréhender en raison de la multiplicité de ses sources, de la diversité de ses formats de données, et de la dispersion de ses plateformes. Il constitue néanmoins le lieu principal des prises de parole publiques actuelles. Dans ce contexte, le Medialab de Sciences Po Paris a initié une réflexion autour de la nécessité d'outils adaptés aux humanités numériques pour l'analyse de corpus de grande envergure. Franck Ghitalla, chercheur en Sciences de l'Information et de la Communication à l'UTC de Compiègne, a contribué à l'élaboration d'une méthodologie pour l'étude des prises de parole sur le Web. Son travail s'inscrit dans la sociologie des controverses, une orientation de recherche notamment portée par Bruno Latour et Michel Callon au sein du CSI de l'École des Mines. Même s'il n'est pas directement affilié au Medialab, l'impact de Franck Ghitalla est notable. Il a constitué un réseau de développeurs web qui oscillent entre le monde académique et l'entrepreneuriat, et a contribué à intégrer la sociologie structurale de l'analyse des réseaux sociaux dans l'étude du Web. Le logiciel Gephi, développé par les frères Jacomy, est devenu un outil couramment utilisé pour l'analyse des réseaux sociaux. Ce logiciel est reconnu pour sa capacité à analyser des ensembles de données complexes et est largement adopté dans des contextes académiques. Dans cette optique, le programme My Web Intelligence apparaît comme une étape supplémentaire dans le développement des outils et des méthodes qui forment la base de l'intelligence numérique et des études digitales. Ce programme cherche à fournir des solutions méthodologiques pour l'analyse du discours en ligne.

Lorsqu'Amar Lakel intègre l'Université de Bordeaux en 2006 en tant que chercheur en Sciences de l'Information et de la Communication, il clôt un cycle de recherches centrées sur les mutations de la gouvernance publique à l'ère d'Internet. Sa thèse postule que la nature ouverte et complexe d'Internet remet en question les fondements de la gouvernance étatique telle qu'elle s'est construite depuis le XIV<sup>e</sup> siècle. Confronté à une obligation paradoxale d'innovation pour maintenir sa fonction structurante, l'État doit réviser ses modalités de gouvernance, non seulement pour réguler la société, mais également pour initier des changements sociaux et des innovations. Après une collaboration sur le premier projet ANR en Sciences Humaines et Sociales sur la gouvernance d'Internet sous la direction de Françoise Massit-Folléa à la Fondation de la Maison des Sciences de l'Homme (2004-2006), Amar Lakel se lance dans un projet de recherche sur l'observatoire du web administratif français dès son arrivée à Bordeaux. Dans sa quête d'un outil adapté à la cartographie du territoire numérique, il s'intéresse initialement au logiciel Tulip, développé par le LABRI de l'Université de Bordeaux. Cependant, Tulip, conçu principalement pour des informaticiens dans un contexte de big data, ne répond pas, à l'époque, totalement à ses besoins. Il assiste donc à la première présentation du logiciel Gephi par Mathieu Jacomy, marquant un moment précurseur pour le

Medialab. L'occasion de concrétiser ses recherches se présente lorsque la Fing, en charge du projet "Rive Droite Numérique" de Bordeaux pour le GIP GPV en 2008, sollicite Lakel pour établir une méthodologie d'étude et de gouvernance numérique des territoires. Il propose alors de collaborer avec Linkfluence pour la cartographie des réseaux sociaux. Bien que le projet constitue une avancée, il ne satisfait pas complètement aux critères d'une démarche scientifique rigoureuse. Répondant ensuite à un appel de la société SFR sur l'analyse des controverses socio-techniques en ligne et ayant mené un audit sur les meilleures pratiques de veille stratégique en 2009, Lakel recentre ses recherches sur les enjeux épistémologiques et méthodologiques de l'analyse des corpus numériques. Au croisement de la théorie des graphes, du text mining et des théories de l'influence, il lui faudra quatre années pour mettre au point un dispositif d'analyse et de territorialisation des corpus numériques, ainsi que de cartographie des controverses. Cette trajectoire de recherche illustre la complexité et les défis associés à l'élaboration de méthodologies rigoureuses dans le champ de l'intelligence numérique et des humanités digitales. Elle met également en lumière l'importance de l'interdisciplinarité et de la collaboration entre différentes sphères du savoir pour adresser les questions multifacettes liées à la gouvernance, à l'innovation et à la compréhension des espaces numériques.

La transformation radicale de l'architecture du Web et les avancées en matière de théorisation de l'activité en ligne ont instauré un tournant épistémologique majeur dans la compréhension de la gouvernance numérique et des discours publics sur Internet. Trois axes conceptuels majeurs catalysent cette transformation :

1. Les Infomédiaires en tant qu'Oracles : Contrairement à la représentation populiste d'un Web sauvage et obscur, le Web 2.0 et les applications sociales ont orchestré une hégémonie d'intelligence collective. Ces infomédiaires, que le projet My Web Intelligence qualifie d'"oracles", sont des actants cruciaux dans cette dynamique. Ils servent de portails d'accès pour naviguer et catégoriser cet espace numérique, en s'appuyant sur des moteurs de recherche, des bookmarks collaboratifs, des réseaux sociaux et l'Open Data. Ils constituent ainsi les mécanismes heuristiques permettant de sonder les abysses de l'internet interconnecté et d'en extraire des morceaux de discours public via des crawlers.
2. Structuration Sociale de l'Hypertexte : Les théories de l'hypertextualité et divers algorithmes sophistiqués (comme la théorie des graphes, le deep learning, et les analyses de régression et multifactorielles) ont mis en évidence que l'interaction numérique est loin d'être chaotique; elle est, en fait, structurée et socialisée. L'exégèse de ces interactions peut révéler non seulement des profils d'acteurs et des réseaux de relations, mais également des dynamiques de contenu, incluant les controverses en ligne.
3. Automatisation de l'Analyse Textuelle : L'abondance des corpus disponibles a longtemps été perçue comme une entrave à la compréhension et à l'analyse du discours en ligne. Toutefois, avec l'évolution de l'IA dans le champ du text mining et de la vectorisation du contenu, il est

désormais possible de procéder à des classifications ontologiquement avancées. Cette mutation permet d'explorer des méta-textes à partir de vastes corpus de données et de naviguer dans ces derniers grâce à des techniques de data visualisation.

La prémisse que le Web est un espace limité, socialement structuré et intelligible, gagne en crédibilité et se métamorphose en un défi technologique qui sert de fondement au projet My Web Intelligence. Ce projet aspire à fournir aux acteurs économiques, institutionnels et sociaux une cartographie axiologique du Web, se constituant en une plate-forme robuste de sondes pour analyser et visualiser le discours public en ligne. Les répercussions de cette plate-forme sont considérables et touchent toute entité utilisant le Web pour l'analyse du discours public. Ce que cette évolution met en exergue, c'est un basculement paradigmatique dans la façon dont nous conceptualisons et abordons le Web en tant qu'arène discursive. Cela engendre des problématiques cruciales, non seulement au niveau des humanités numériques, mais également en ce qui concerne les études sur la gouvernance, la communication, et l'impact sociétal du Web. Ce n'est pas simplement une question de technologie plus sophistiquée, mais bien d'une reconfiguration fondamentale dans notre compréhension des modalités de l'interaction humaine et institutionnelle à l'ère numérique.

## **II.2 – De l'enquête des controverses en études digitales : enjeux des digital methods**

Le projet My Web Intelligence ne se contente pas de poser des briques technologiques et méthodologiques ; il instaure également un paradigme éthique et collaboratif. En optant pour une licence MIT open-source, le projet convoque une logique de fédération pluridisciplinaire. Plus qu'une simple démocratisation des outils d'analyse du web, cette ouverture invite chercheurs, développeurs et experts de l'information à contribuer à une intelligence collective. Ce faisant, il affronte le défi colossal de rendre intelligibles les profusions d'énonciations publiques sur le web. Le web, dans sa structure actuelle, souffre d'une discontinuité qui le rend en partie inintelligible. L'invisibilité du web, son évanescence dans la cacophonie des contenus, en fait un espace public distordu. Les résultats que l'on obtient via des moteurs de recherche sont souvent des unités informationnelles atomisées, dénuées de leur contexte et de leurs conditions de production. Cette décontextualisation ampute l'information de son intégrité, en excluant les éléments socio-linguistiques et stratégiques qui lui donnent du sens. Inspirés par les sociologies des controverses, de l'innovation et de la traduction, nous considérons que la compréhension des archives numériques nécessite un déchiffrement multi-niveaux. Il faut aller au-delà du document lui-même pour inclure le métadiscours qui l'entoure et les locuteurs qui le produisent. Cette approche holistique permet de remettre en perspective le contexte socio-linguistique de l'information et de redécouvrir la dynamique sociale qui constitue à la fois le milieu de sa production et le terrain de sa réception. La forme canonique de la "liste" de résultats, dépourvue de tout contexte, a jusqu'à présent écarté toute forme d'intelligence émise par les interlocuteurs. Elle rend



obsolète tout effort de discernement des jeux stratégiques qui sont en jeu dans ces prises de paroles. En d'autres termes, la "liste" ignore l'écosystème dynamique de production et de réception des informations. Elle inhibe toute possibilité d'engager une herméneutique du web basée sur une intelligence collective. En synthèse, le projet My Web Intelligence s'emploie à combler ce déficit d'intelligibilité en proposant un cadre d'analyse sophistiqué et multi-niveaux qui englobe non seulement les documents, mais également les agents et les stratégies en jeu. Cette ambition s'inscrit dans une démarche plus large qui vise à réintroduire de l'intelligence collective dans la navigation et la compréhension du web. Ainsi, My Web Intelligence ne sert pas uniquement d'outil d'analyse mais devient une plateforme collaborative pour l'éclairage et la contextualisation des discours numériques.

La plateforme My Web Intelligence se positionne comme un artefact technologique et méthodologique dans le domaine des études digitales, tout particulièrement en ce qui concerne l'exploration et l'analyse des controverses en ligne. Cette plateforme, hébergée au sein du MICA et s'exécutant dans un environnement Python avec une base de données SQLite, symbolise une forme de cristallisation des avancées techniques et conceptuelles que la discipline a connues jusqu'à ce jour.

1. Gestionnaire de Territoire : Ce composant fait office de gestionnaire de projet. Il permet à un utilisateur d'initialiser une exploration web en définissant des requêtes basées sur des mots-clés. Ces dernières servent de fondement pour l'extraction des données et leur représentation visuelle. À terme, ce gestionnaire envisage d'intégrer des fonctionnalités pour la gestion collaborative et la mise en visibilité publique des projets.
2. Extracteur de Données Numériques et Services d'Annotations Associées : Ce module est chargé de la constitution des archives numériques. Il intègre un serveur HTTP natif destiné à absorber les ressources numériques, ce qui témoigne de la modularité et de l'évolutivité de l'architecture.
3. Visualisation des Grandes Masses de Données : La plateforme est équipée d'un tableau de bord (Dashboard) permettant de gérer d'importants corpus de données via une série de marqueurs. Plus significativement, elle génère des cartographies relationnelles qui, à travers l'analyse des graphes, mettent en lumière la structure hiérarchisée des données.

À l'issue de la phase de prototypage, My Web Intelligence aura apporté des solutions à une variété de défis techniques et scientifiques, ouvrant la voie à de nouvelles avenues de recherche et d'application. Le passage du prototype à un service stable et professionnel constitue cependant une série de défis inédits. Parmi eux, notons la nécessité de résoudre des problématiques algorithmiques complexes ("carrés algorithmiques bloquants"), de gérer les défis inhérents au Big Data, et de créer une infrastructure robuste capable de supporter des accès multi-utilisateurs. Ainsi, le prototype actuel sert plus comme un appel à un programme de recherche continu qu'une solution définitive. La matérialisation de My Web Intelligence en tant qu'instrument d'analyse des controverses dans le

domaine digital exige une réflexion épistémologique profonde. Il ne s'agit pas seulement d'une technologie en soi, mais aussi d'une méthodologie qui façonne et est façonnée par les paradigmes scientifiques qui régissent les études digitales. Ce sont ces interactions complexes et réciproques entre la technique et l'épistémologie qui confèrent à My Web Intelligence son caractère innovant et, par conséquent, son potentiel pour remodeler le paysage des études digitales et des humanités numériques.

Dans l'espace numérique contemporain, où la quête d'information se réduit souvent à des requêtes parfois stériles dans des moteurs de recherche, My Web Intelligence ambitionnent d'offrir une alternative. Ce ne serait pas tant une solution déjà aboutie, mais plutôt un programme en devenir, une tentative algorithmique de complexifier ce que l'on simplifie souvent à outrance. Dans un idéal de renversement épistémologique, cette plateforme souhaiterait convertir la requête instantanée en un projet de recherche en évolution constante. L'utilisateur serait encouragé à envisager sa quête informationnelle non comme une transaction ponctuelle, mais comme un index structuré et dynamique, évoluant au gré des interactions et des choix de pertinence. Alors que les moteurs de recherche traditionnels reposent sur une rationalité limitée, la démarche prospective de My Web Intelligence viserait à créer une ontologie du savoir nettement plus nuancée. L'ambition serait de parvenir, via un apprentissage machine assisté, à un écosystème où le système algorithmique s'adapte et évolue, captant les nuances humaines au fur et à mesure. On pourrait y voir moins une prestation de service qu'une sorte de dispositif archéologique virtuel, qui non seulement cartographierait les données mais permettrait aussi des annotations, des réflexions et des réajustements. L'architecture fonctionnelle de la plateforme, toujours en amélioration, chercherait à intégrer un éventail de fonctionnalités -comptes utilisateurs personnalisés, fonction de mise à jour des données, options de travail collaboratif- tout en poursuivant l'objectif d'une mémoire centralisée. My Web Intelligence ne se limiterait pas à l'oracle unique que représente souvent Google, mais se positionnerait comme un agrégateur d'oracles multiples, dans une tentative de démocratiser le paysage informationnel. L'intelligence algorithmique de la plateforme aspirerait à explorer des couches plus profondes de l'Internet, à filtrer le superflu et à prioriser l'information. Ici, l'algorithmie se verrait appuyée par des mécanismes d'approbation humaine, visant à créer un système plus synergique. Chaque document deviendrait une entité analysable, pouvant être annotée pour enrichir l'analyse socio-linguistique de la situation de communication. Ce serait un jeu à somme positive : l'intelligence collective pourrait se former à mesure que chaque utilisateur enrichit la base de données. Cependant, il faudrait considérer My Web Intelligence comme un chantier en cours plutôt qu'une réussite confirmée. Il s'agirait d'une tentative de théorisation multidimensionnelle de la recherche sur Internet, une infrastructure en développement pour naviguer et, éventuellement, gouverner le labyrinthe informationnel qu'est le web contemporain. La plateforme ambitionnerait de transcender le dualisme simpliste entre requête et réponse, aspirant à établir un dialogue symbiotique et évolutif entre utilisateur, algorithme et réseau.

En conclusion, l'enjeu majeur de l'exploration web, comme nous l'avons précédemment noté, repose sur la capacité à annoter, classifier et interpréter un corpus d'une envergure quasi illimitée. Il convient toutefois de rappeler que les projets visant une intelligence véritablement performative se concentrent généralement non sur "le web" dans son intégralité, mais sur des sous-ensembles spécifiques de celui-ci. Plus ce sous-ensemble est volumineux, plus les ressources humaines nécessaires à son étude sont considérables. L'idéal vers lequel aspire My Web Intelligence est donc bien celui de la qualification automatique des données les plus pertinentes. Bien que l'automatisation totale soit une chimère dans le domaine de l'intelligence artificielle, des avancées notables dans l'analyse automatique des documents permettent d'envisager une annotation de contenu de page entièrement automatisée à court terme. Cependant, la tâche qui s'avère particulièrement ardue est celle de la qualification du média au sein de son environnement social et discursif. D'importantes avancées en matière d'automatisation sont certes envisageables, mais c'est dans le domaine de la gestion de la connaissance collective, notamment à travers le crowdsourcing de l'annotation humaine, que le potentiel est véritablement colossal. L'architecture idéale que pourrait adopter My Web Intelligence permettrait la formation de collectifs d'experts spécialisés dans différents domaines ou secteurs informationnels. Ces experts pourraient partager leurs ressources dans la qualification et l'annotation de sous-ensembles sectoriels du web. En ce sens, la communauté des utilisateurs de My Web Intelligence pourrait rapidement être sollicitée pour collaborer à la construction d'indices sectoriels de qualification du web, dans le cadre de projets en open data. À long terme, cette "mémoire partagée" cumulative constituerait un capital immatériel inestimable. Elle fournirait aux nouveaux utilisateurs un système d'indexation sectorielle déjà riche et nuancé, minimisant ainsi la friction souvent associée à la recherche d'information sur le web. Ce serait un pas vers l'édification d'une infrastructure d'information plus démocratique et inclusive, tout en optimisant l'efficacité et la pertinence de la recherche informationnelle.

### **II.3 – L'analyse comme re-fabrication du contexte stratégique de l'acteur réseau**

À l'aube du XXI<sup>e</sup> siècle, nous nous trouvons à un carrefour épistémologique similaire à celui que la sociologie a rencontré au début du siècle précédent. Internet, cette vaste agora numérique, agit comme un immense réceptacle d'opinions, un recensement a priori de toutes les réponses à toutes les questions imaginables, avant même leur articulation formelle. Collecter, analyser et interpréter ces données de manière à contribuer à une intelligence sociale renouvelée constitue un défi herméneutique et méthodologique. Le prototype My Web Intelligence incarne une réponse préliminaire à cette question, mais il sert surtout de manifeste pour prendre au sérieux les enjeux de l'intelligence numérique. Les sciences sociales et humaines ont toujours été caractérisées par une démarche de réduction de la complexité. Ce n'est pas une limitation, mais plutôt une condition sine qua non de leur

exercice. Le but est de distiller la multitude d'événements sociaux en un ensemble saisissable de rôles, de comportements, de structures collectives et de dynamiques. Cette réduction est non seulement constitutive, mais elle est également indispensable pour toute forme de gouvernance. Cependant, il est crucial de ne pas perdre de vue cette réduction lorsqu'on examine les phénomènes à une échelle microscopique. Chaque individu, ou plutôt chaque instance de prise de parole sur Internet, est approximé par ces réductions conceptuelles. Il existe une dualité inhérente à chaque individu : une part qui le subsume à un objet commun, et une autre qui demeure irréductible, originale et unique. Les concepts sociologiques agissent alors comme des coordonnées dans un espace tridimensionnel, permettant de situer le document numérique dans un univers médiatique, produit par un comportement discursif type et activant une réalité construite. Mais en adoptant une perspective macroscopique, l'espace numérique peut être envisagé comme étant structuré selon trois réalités distinctes, bien que la question d'une analyse multidimensionnelle demeure ouverte. Cette structuration tridimensionnelle offre un cadre pour une analyse plus nuancée et complexe, ouvrant la voie à de futures recherches dans le domaine des humanités numériques. Nous sommes à un moment charnière, nécessitant une mobilisation et des investissements significatifs en recherche et développement pour intégrer pleinement l'espace public numérique. Les enjeux sont colossaux, mais les possibilités sont tout aussi vastes. Le prototype "My Web Intelligence" n'est qu'un premier pas vers la réalisation de cette vision, un programme qui aspire à transformer notre compréhension de l'intelligence sociale à l'ère numérique. Ce projet aspire à contribuer à la formalisation d'outils et de méthodologies qui permettront aux sciences sociales de naviguer et de comprendre la complexité du monde numérique, tout en respectant les principes épistémologiques qui ont guidé la discipline depuis sa genèse. My Web Intelligence, c'est aussi une méthode avec ses processus qui s'inscrivent dans la volonté de scientificité et donc de reproductibilité des analyses en sociologie numérique. Une fois l'extraction et la qualification des données d'une controverse achevées, My Web Intelligence donne accès à un corpus nettoyé qui permet de pouvoir mettre en place un ensemble de traitements d'analyse et de traitements des données pour tirer véritablement une compréhension de l'économie de la discussion en ligne.

L'extraction de données sur le web pour la constitution de corpus représente un enjeu méthodologique premier et fondamental, en un mot fondateur, dans la recherche en sociologie numérique. Dans une recherche sur les influenceurs en information de santé sur Internet, la requête détermine l'objet sémantique étudié. Par exemple, la requête "asthme (enfant OR adolescent OR jeune OR bébé OR nourrisson OR infantile)" soumise aux moteurs de recherche Google et Bing, aboutit à l'extraction de 1235 pages de premier niveau (depth=0), regroupées ensuite en un second set de 846 domaines d'expression. Cette démarche visait à simuler le comportement d'un internaute français moyen visant à s'informer sur une maladie courante chez l'enfant. Étant donné que plus de 90 % des requêtes en France sont effectuées via Google, l'usage de mots simples et courants détermine le champ à explorer (ici celui de l'éducation à la maladie). Les termes de la requête ont été choisis pour être

volontairement neutres et courts afin de minimiser les biais d'échantillonnage souvent introduits par un vocabulaire socialement situé. En effet, le choix d'un vocabulaire spécifique—qu'il soit technique, scientifique, familier ou affectif— constitue un choix d'échantillonnage qui sur-représente nécessairement les populations habituées à utiliser ces mots-clés. Il est crucial de comprendre que le Web signifiant pour un utilisateur de ces moteurs de recherche est un tout petit échantillon du web limité, structuré selon trois axes principaux :

1. Formellement Structuré : Les algorithmes de ranking des moteurs de recherche imposent une hiérarchie dans les résultats, influençant ainsi la visibilité et l'accessibilité des informations.
2. Socialement Structuré : Les stratégies marketing des locuteurs sur le web jouent également un rôle dans la structuration des résultats, notamment à travers le SEO (Search Engine Optimization).
3. Intentionnellement Structuré : La requête de l'utilisateur, qui est le point de départ de cette interaction, impose ses propres limites en fonction des termes choisis.

Il est certes courant d'affirmer que l'internaute "surfe sur le web", mais cette expression simpliste escamote la complexité de l'expérience de navigation en ligne. En effet, la trajectoire de l'utilisateur sur Internet est façonnée par des infomédiaires tels que les moteurs de recherche (notamment Google et Bing), les plateformes de médias sociaux (Facebook, Twitter, etc.), les forums de discussion, et même les fournisseurs de services de messagerie électronique. Ces entités agissent comme des filtres médiatiques qui sont régis par une politique éditoriale particulière, laquelle est elle-même une réification algorithmique de règles, de sélections communautaires, et d'un profil comportemental historisé de l'utilisateur. Ce trio de critères conditionne la constitution d'un corpus numérique limité à un certain champ sémantique, un fait qui pose des questions épistémologiques non négligeables relatives à la validité et la représentativité des données extraites. La tâche d'extraction de données pour la constitution de corpus web n'est en aucun cas une opération neutre; elle est imprégnée d'une multitude de considérations méthodologiques, épistémologiques et éthiques. La reconnaissance de ces biais et limitations s'avère essentielle pour la rigueur scientifique en sociologie numérique. Par conséquent, il est capital de conceptualiser des méthodologies réflexives qui reconnaissent ces imbrications complexes, tout en cherchant à atténuer leur incidence sur l'intégrité et la qualité de la recherche. Bien souvent, les études qui reposent sur des corpus numériques tendent à négliger cette posture épistémologique réflexive. Or, une telle omission revêt des implications significatives. Elle passe sous silence les différentes strates de conditionnement qui influent non seulement sur la collecte, mais également sur l'herméneutique des données au sein de l'écosystème numérique. A chaque étape de la démarche scientifique en sociologie numérique, depuis l'élaboration de la problématique jusqu'à l'interprétation des données, il est fondamental de dévoiler les mécanismes épistémologiques et méthodologiques sous-jacents. C'est par cette explicitation que l'on peut prétendre à une certaine

rigueur scientifique, une rigueur qui prend en compte la complexité de la médiation technologique dans la production et la circulation du savoir.

La seconde phase de la recherche réside dans la structuration et l'enrichissement de données par annotation des données. Cette étape permet de qualifier et de contextualiser les données par des métadonnées, fournissant ainsi une base pour l'analyse subséquente.

1. La première étape consiste en une extraction automatique des données au niveau de la page web. Cette extraction récupère le code HTML complet, le titre (title), l'URL et la description contenue dans les balises méta. Des algorithmes sont ensuite utilisés pour détecter le contenu principal du document (readable), ainsi que les URL de liens (urls), d'images (imgs) et de vidéos (videos) présents dans ce contenu. En cas d'extraction par crawling à plusieurs niveaux de profondeur, une variable 'depth' indique si le document a été obtenu au démarrage (niveau 0) ou par extraction subséquente (niveau n).
2. Une deuxième couche d'extraction est effectuée via des services web par API. Cette étape permet de mesurer la résonance sociale de l'URL à travers différentes variables, telles que le nombre de "likes" et de "shares" sur Facebook, l'autorité globale des pages et des domaines par des services comme Moz ou Semrush, etc.
3. La troisième étape implique une annotation humaine pour contextualiser davantage le document. Trois variables sont utilisées à cette fin : la date de publication (publication\_date) sur le moteur de recherche Google, le type de média (media\_type) et le type d'émetteur (emitter\_type). Cette dernière variable est conçue pour être ouverte afin de capturer une gamme variée de types d'émetteurs.
4. Un premier travail de TAL est effectué par un algorithme de lemmatisation pour qualifier le contenu par les titres et les contenus de la page readable. Ce traitement permet de déterminer la pertinence de la page au regard du dictionnaire de mots clés.
5. Dans cette recherche, trois variables ont été définies pour ancrer la position de l'énonciateur dans un espace matérialisé. Les *types d'acteurs* sont un champ ouvert et qualitatif qui vise à appréhender le statut socioprofessionnel du locuteur en essayant d'être guidé à la fois par la source de légitimité (acteurs de la santé, acteurs publics, journalistes, entreprises à but lucratif, usagers et patients et des professions alternatives) et son mode d'institutionnalisation (individus, associations, entreprises et institutions à but non lucratif). Les *types de médias* visent à classer les types de dispositifs de communication comme autant de forme institutionnalisant un contrat de communication par des usages partagés (Ghiglione, 1986; Perriault, 1989; Flichy, 1995; Chaudiron et Ihadjadene, 2010).

Cette méthode d'annotation sert à ajouter des variables à un ensemble de données, à différencier du processus de recodage, qui consiste à produire de nouvelles variables à partir de variables existantes. Elle offre une base pour une analyse de contenu en isolant le discours principal de la partie éditoriale de la page web<sup>3</sup>.

La seconde étape de notre méthodologie repose sur l'analyse des réseaux, en utilisant spécifiquement la théorie des graphes et l'analyse structurale des réseaux pour générer des cartographies des médias impliqués dans la controverse. Cette démarche vise à aller au-delà des mots pour identifier les locuteurs qui opèrent ces médias et qui sont engagés dans un espace public numérique. Il est impératif de qualifier ces médias selon leur nature sociale et leurs comportements éditoriaux. Plus important encore, il faut révéler, à travers la structure de leurs citations qualifiées, le contexte d'alliance et d'adversité qu'ils tissent. Cette analyse s'appuie sur le principe que les citations et les références d'un média révèlent sa position sociale et idéologique ("Dis-moi qui tu cites, et je te dirai qui tu es"). Une vision globale et structurale des acteurs permet de déceler non seulement la structure des alliances et des oppositions, mais aussi les communautés d'intérêts idéologiques. Elle permet également de situer chaque média selon un rôle social spécifique dans le débat et au sein de sa communauté. Ces rôles peuvent être ceux de leader d'opinion, de vigie, de marginal sécant, de bridge, etc. L'analyse des graphes sert également à comprendre la structuration argumentaire latente. Les prises de position contribuent à la construction d'une réalité collective. L'analyse de la coprésence d'objets du discours, comme les mots-clés, révèle une structure de représentation collective. Les variables topologiques des graphes, telles que le "betweenness centrality" et le "Page Rank local", permettent de comprendre le rôle de chaque concept dans une stratégie argumentaire et d'identifier les glissements notionnels. L'ensemble des algorithmes d'analyse structurale des réseaux est implémenté dans le logiciel Gephi. Pour chaque URL, des métriques comme les degrés orientés (indegree et outdegree) et pondérés (weighted degree, weighted indegree, weighted outdegree) sont calculés. L'algorithme de détection/qualification des communautés par "Modularity Class" permet de déceler les communautés de citations du réseau. D'autres métriques, telles que l'Authority Rank et le Hub Rank, sont utilisées pour évaluer et classer les domaines selon leur autorité et leur rôle d'infomédiaires. En outre, des calculs de "closeness centrality" et de "harmonic closeness centrality" sont effectués pour identifier les nœuds qui sont au cœur du réseau. L'analyse en graphe des arguments, une fois datés, permet d'historiciser la construction des représentations collectives et d'en comprendre la dynamique. Cette étape d'analyse des réseaux sociaux est donc cruciale pour examiner les structures complexes et les dynamiques de pouvoir qui sous-tendent les controverses publiques dans l'espace numérique.

---

<sup>3</sup> Il faut rappeler que le document Web est un construit collectif qui engage l'auteur, l'éditeur, le développeur point c'est 3 statuts étant eux-mêmes des collectifs répartis dans l'espace et dans le temps point ainsi le document web et le fruit d'une négociation et d'une construction complexe. Nous avons pris comme parti de nous intéresser et donc d'isoler l'auteur du contenu du processus d'éditorialisation.

La troisième étape de notre méthodologie vise à opérationnaliser l'Actor-Network Theory (Latour, 2007; 2010; 2014) dans le contexte des Digital Media Studies (Barat, 2017; Boullier, 2016). Cette étape est cruciale pour l'analyse des prises de position sur le web, car elle permet une exploration systématique des algorithmes de traitement automatique de la langue (NLP). Elle s'inscrit dans une démarche multidimensionnelle qui englobe les comportements éditoriaux, les identités socialement situées et la mise en réseau des acteurs par leurs stratégies de citation. Une phase de "Text Mining et Prétraitement" constitue un jalon crucial dans notre méthodologie de recherche. Elle s'inscrit dans le cadre plus large de l'analyse des prises de parole dans l'espace numérique. L'objectif principal de cette étape est de transformer le corpus textuel en une forme structurée et analysable, tout en préservant les nuances et les complexités inhérentes aux discours individuels et collectifs. Pour ce faire, nous utilisons la librairie `tm` de R, un outil puissant pour le text mining. Le prétraitement est une étape préliminaire indispensable qui prépare les données pour des analyses plus poussées. Dans notre cas, le contenu textuel, déjà "readabilisé" (c'est-à-dire transformé en une forme plus lisible et structurée), subit une série de transformations. Ces transformations incluent la lemmatisation, qui réduit les mots à leur forme canonique. Par exemple, les mots "cours", "courrir" et "course" seraient tous réduits à leur lemme "course". La lemmatisation permet de créer un dictionnaire global des lemmes retenues par l'analyste, qui sert ensuite de base à la vectorisation du texte. Dans notre cas, cette vectorisation s'est faite dans un espace à 790 dimensions, chaque dimension correspondant à un lemme du dictionnaire. Chaque document du corpus est alors transformé en un vecteur dans cet espace à 790 dimensions. Ce vecteur, ou "empreinte numérique", représente une caractérisation multidimensionnelle du contenu du document. La création de cette empreinte numérique est fondamentale pour les étapes ultérieures de l'analyse, notamment la Classification Ascendante Hiérarchique (CAH) et l'Analyse des Composantes Principales (ACP). Elle permet non seulement une réduction dimensionnelle efficace mais aussi une représentation plus nuancée des prises de position. Vient ensuite la phase de "Classification et Thématisation" qui représente une étape analytique qui succède au prétraitement et à la vectorisation des données textuelles. L'objectif de cette phase est de catégoriser et de structurer le corpus textuel en fonction de thèmes ou de sujets dominants, tout en préservant la complexité et la diversité des prises de position. Cette étape est indispensable pour une analyse fine des dynamiques discursives et des structures argumentatives au sein de l'espace numérique. Pour atteindre cet objectif, nous avons employé une Classification Ascendante Hiérarchique (CAH), une technique de clustering qui permet de regrouper les documents en fonction de leur similarité dans l'espace vectoriel à 790 dimensions. Cette CAH est renforcée par l'emploi de la méthode des K-Means, une autre technique de clustering qui optimise la partition des données en minimisant la variance intra-cluster et en maximisant la variance inter-cluster. L'application de ces méthodes a conduit à la formation de sept clusters distincts, ou thèmes majeurs, qui sont statistiquement discriminés. Ces thèmes représentent les axes discursifs dominants au sein du corpus et offrent une grille d'analyse pour comprendre les différentes prises de position. Chaque thème peut être considéré comme un ensemble cohérent de discours, d'arguments, et



de points de vue qui partagent une certaine logique interne et qui interagissent de manière spécifique avec les autres thèmes. La classification et la thématisation ne sont pas de simples exercices taxonomiques; elles constituent plutôt des outils épistémologiques qui permettent de naviguer dans la complexité des données. Elles offrent des perspectives analytiques pour étudier les dynamiques de pouvoir, les structures d'alliance et d'opposition, et les processus de légitimation au sein de l'espace discursif. En somme, cette phase permet de passer d'une masse de données textuelles à une structure organisée et interprétable, ouvrant la voie à des analyses plus nuancées et plus profondes des prises de position dans l'espace numérique. L'application du NLP et du text mining dans cette étape permet non seulement de thématiser le contenu mais aussi d'opérationnaliser la notion de prise de position dans l'espace numérique. Elle offre une méthodologie robuste pour l'analyse des discours en ligne, permettant de cartographier les nuances et les complexités des prises de position dans des débats publics numériques.

# III – Etude comparée sur les influenceurs en information de santé sur l'internet : le cas du diabète et de l'asthme chez l'enfant.

C'est à partir d'un corpus richement annoté que peut s'engager alors l'analyse et l'interprétation des énonciations sur Internet. L'analyse multivariée joue un rôle significatif dans notre méthodologie, en particulier pour l'identification des profils comportementaux des locuteurs. Cette technique statistique permet de dégager des cooccurrences à partir de plusieurs variables, telles que la nature sociale des locuteurs, les sujets abordés, les arguments utilisés, les dates de publication et les types de médias utilisés. Ces co-occurrences sont obtenues par de nombreuses méthodes de clusterisation (ACP, Kmean, etc.). L'analyse multivariée permet non seulement de quantifier mais aussi de qualifier les prises de position types. Elle offre la possibilité de comprendre leur évolution dans le temps. Par exemple, derrière un grand nombre de messages, il est souvent possible de dégager 3 ou 4 positions majeures qui dominent l'espace public discursif. En outre, rarement plus d'une dizaine de ces positions sont responsables de la production de la quasi-totalité des discours en ligne. L'analyse multivariée contribue à une compréhension plus synthétique de la structuration de la controverse au fil du temps. Si l'analyse structurale avait pour but de "socialiser" les médias, la réduction du corpus à quelques profils comportementaux quantifiés permet d'obtenir une vision plus claire de la dynamique de la controverse. Cette clarté est particulièrement utile pour les chercheurs et les analystes qui cherchent à comprendre les mécanismes sous-jacents qui façonnent l'opinion publique en ligne. L'analyse multivariée sert donc comme un outil complémentaire à notre méthode d'annotation, permettant une analyse plus profonde et plus nuancée des données. Elle facilite la décomposition de l'espace discursif en profils comportementaux, enrichissant ainsi notre compréhension des dynamiques sociales et discursives qui caractérisent l'espace public numérique.

## III.1 - Quantification du corpus des énoncés numériques

L'étude de cas cherche à élaborer une cartographie du paysage communicationnel dans le secteur de l'information de santé. Elle vise à décomposer la typologie des locuteurs et à analyser leurs choix stratégiques en matière de plateformes communicationnelles. L'objectif ultime est de déchiffrer les dynamiques discursives et les structures de pouvoir qui influencent ce domaine. La méthodologie adoptée pour cette étude est pluridisciplinaire, combinant des approches quantitatives et qualitatives

pour obtenir une vue d'ensemble du paysage. Elle s'appuie sur le set de données annotés pour retenir dans un premier temps des informations à partir des divers sites et plateformes en ligne. En utilisant comme variable le *Type de Locuteur* (Catégorisation des acteurs en fonction de leur position sociale: acteurs de santé, journalistes, patients, entreprises, etc.), le *Type de Plateforme* (Classification des plateformes utilisées pour la communication : sites éditoriaux, institutionnels, réseaux sociaux, etc.) et la *Fréquence de Publication* (Mesure du volume de contenu produit par chaque type de locuteur), on vise à mesurer la stratégie de communication de chaque acteur en évaluant les tactiques employées par les locuteurs pour atteindre leur public cible (stratégie éditoriale, institutionnelle, conversationnelle, etc.). Il nous faut évaluer la répulsion/attraction des acteurs pour certaines stratégies ou plateformes afin de distinguer les différents sous-espaces de communication. Une simple analyse des fréquences couplé à des test du Khi2 permettent de détecter des sur et sous-représentations des acteurs en fonction de leur choix de plateforme. Une analyse factorielle, une Classification Ascendante Hiérarchique (CAH) et le K-Means viennent alors renforcer l'analyse pour réduire la complexité des données et identifier les variables latentes qui expliquent les variations observées.

La prédominance des acteurs de santé et des journalistes dans le discours sur l'information de santé est manifeste, occupant respectivement 60,8% des sites et 67,9% des pages. Cette hégémonie discursive n'est pas anodine; elle suggère une concentration du capital symbolique et épistémique entre les mains d'acteurs représentant un très haut degré d'institutionnalisation. Leur influence dans le champ de l'information de santé peut avoir des implications majeures, notamment en termes de légitimation de certaines formes de savoir médical et de marginalisation d'autres voix. En effet, les patients et les entreprises émergent comme des acteurs secondaires (mais non négligeables), représentant 13,59% et 13,26% des pages respectivement. Leur présence modérée indique certes une certaine démocratisation de l'espace discursif mais tout de même limitée. Les acteurs publics hors du domaine de la santé et les alternativistes (prônant une position en opposition avec le secteur de la santé) sont largement marginalisés, avec seulement 3,15% et 2,10% des pages alors même qu'ils nourrissent les angoisses journalistiques sur les dangers de l'internet. Cette marginalisation est révélatrice des dynamiques de pouvoir qui structurent le discours public dans ce domaine, et suggère une homogénéisation du discours autour des paradigmes médicaux et journalistiques dominants. Du côté des plateformes investies, les sites éditoriaux dominent le paysage avec une proportion de 31,8%, suivis des sites institutionnels à 25,8%. Cette prédominance des sites éditoriaux et institutionnels indique une stratégie de communication axée sur la légitimation et la formalisation du savoir. Les plateformes Web 2.0, y compris les réseaux sociaux, les forums et les blogs, représentent 17%, ce qui suggère une ouverture vers des formes plus interactives et démocratiques de communication, bien que ces dernières restent minoritaires. Si cette sous-représentation est tout à fait logique quand on accède par le web grâce au moteur de recherche qui vise à faire prédominer les productions éditoriales plutôt que les sites conversationnels, nous sommes néanmoins surpris de voir que les annuaires et les sites

de e-commerce sont nettement peu-représentés, avec seulement 7,3% et 3,9% respectivement tordant le coup à un fantasme d'un internet marchand qui viserait à capitalisé le chant de la santé. Cette sous-représentation pourrait indiquer une moindre commercialisation du discours sur la santé, ou du moins une commercialisation qui s'opère par d'autres canaux. L'utilisation du test du Khi2 a permis une analyse croisée des résidus de Pearson, mettant en lumière des sur et sous-représentations des acteurs en fonction de leur choix de plateforme. Les acteurs de santé et les institutions publiques se cantonnent à une communication institutionnelle, tout en évitant les stratégies de contenu et les réseaux sociaux. À l'opposé, les journalistes et les patients s'éloignent de cette approche corporate. Les journalistes sont particulièrement attirés par les sites éditoriaux, étant les principaux producteurs de contenus périodiques et actualisés. Les patients et les alternativistes, quant à eux, privilégient une stratégie conversationnelle sur les médias sociaux. Ces résultats, tout en étant quantitatifs, offrent une riche matière à interprétation et ouvrent des pistes de réflexion sur les dynamiques de pouvoir, les stratégies de légitimation et les formes de résistance dans le champ de l'information de santé.

Le chapitre qui suit se consacre à la thématisation des contenus relatifs à l'asthme chez l'enfant. Cette exploration s'appuie sur une méthodologie hybride, combinant des techniques de traitement automatique du langage naturel (NLP) avec une qualification humaine. Cette approche dualiste permet une analyse plus nuancée et contextualisée des discours, tout en capitalisant sur la puissance computationnelle pour traiter de grands volumes de données. Le but est de dégager des axes thématiques majeurs et de comprendre les dynamiques discursives qui structurent ce domaine de l'information de santé. Le thème dominant qui émerge de cette analyse est celui de l'éducation thérapeutique, qui occupe une place prépondérante avec 64% des pages analysées. Ce thème se subdivise en trois axes thématiques distincts mais interconnectés. Le premier axe, qui représente 37% des pages, se concentre sur la description clinique de l'asthme chez l'enfant, abordant des aspects tels que les symptômes, les origines et les conséquences de cette maladie. Le deuxième axe, qui occupe 18,3% des pages, adopte une approche d'accompagnement et de conseil, ciblant principalement les mères et proposant des stratégies pour mieux vivre avec l'asthme de leur enfant. Le troisième axe, plus spécifique, se focalise sur les "bons gestes" à adopter pour le bébé et le nourrisson, et représente 8,7% des pages. Un deuxième bloc de discours, qui occupe 23% des pages, est caractérisé par une charge émotionnelle élevée. Ce bloc aborde les facteurs déclencheurs de crises graves et les risques sérieux encourus par les enfants asthmatiques. Il s'agit d'un discours qui, tout en étant informatif, vise également à alerter et à sensibiliser le public. Un troisième bloc, plus sensationnaliste, représente 10,5% du corpus. Ce bloc englobe des thèmes variés, allant de la mort d'un nourrisson dans un hôpital (1,4% des prises de parole) à des études sur des facteurs environnementaux tels que la pollution automobile, les bactéries intestinales et la présence d'animaux domestiques. Ce bloc semble chercher à capter l'attention par le biais du "scoop" et du sensationnalisme. Enfin, un dernier bloc de discours, qui représente 2,7% des pages, est consacré à la promotion de cures thermales pour les asthmatiques. Ce

bloc est principalement porté par des entreprises dans le secteur de la santé et s'inscrit dans une logique commerciale. Dans l'ensemble, près de 80% du corpus est directement lié à une volonté d'éducation thérapeutique, oscillant entre une approche informative et une approche plus alarmiste. Une infime minorité du corpus est constituée de pages sensationnelles qui semblent chercher à créer un "buzz". Cette distribution thématique et émotionnelle du discours révèle une complexité sous-jacente qui mérite une attention particulière, notamment dans le contexte plus large des dynamiques de pouvoir et de légitimation dans le champ de l'information de santé.

Trois facteurs stratégiques émergent comme déterminants dans la structuration du champ de l'information de santé: l'appartenance au monde de la santé, le statut organisationnel et l'appartenance à l'espace marchand ou public. Ces facteurs ne sont pas simplement des variables indépendantes, mais des axes structurants qui façonnent les dynamiques de légitimation au sein de ce champ. L'appartenance au monde de la santé est l'axe majeur qui influence non seulement la visibilité mais aussi la crédibilité des acteurs. Ce facteur est corroboré par le fait que 32,5% des sites appartiennent à des acteurs de la santé. Le statut organisationnel, quant à lui, est un autre axe structurant, avec 80% des sites appartenant à des organisations. Ce pourcentage élevé suggère une forme de "d'institutionnalisation" forte du discours sur la santé, où les organisations ont plus de poids que les individus, qui ne représentent que 20% des sites. Malgré l'ubiquité et la démocratisation apparente du web, nos données révèlent une persistance remarquable des structures et des hiérarchies traditionnelles dans le champ de l'information de santé sur le web 1.0. Les acteurs, bien que diversifiés, semblent s'inscrire dans des logiques communicationnelles qui sont loin d'être révolutionnaires. L'innovation communicationnelle est, de manière assez paradoxale, inversement proportionnelle à son institutionnalisation. Autrement dit, plus un acteur est institutionnalisé, moins il est susceptible d'innover dans ses pratiques communicationnelles. Les alternativistes, bien que quantitativement négligeables, offrent une fenêtre intéressante sur les limites et les possibilités de résistance au sein de ce champ. Leur marginalité numérique (représentant une "pincée négligeable" dans la prise de parole) suggère une forme d'exclusion systémique, mais elle pose également la question de la légitimité et de la reconnaissance dans des espaces dominés par des acteurs institutionnels fortement filtrés aujourd'hui par les algorithmes des moteurs de recherche. Ce qui frappe le plus dans cette étude est la puissante présence de l'industrie journalistique. Occupant la première position en termes de prise de parole, les journalistes semblent agir comme des gardiens du discours public sur la santé. Leur rôle n'est pas simplement de relayer l'information, mais de la structurer, de la hiérarchiser et, inévitablement, de la légitimer. Il est crucial de nuancer ces observations en considérant les sous-groupes au sein des catégories plus larges. Par exemple, les praticiens de santé associatifs et ceux communiquant à titre individuel montrent des stratégies de communication qui dévient de la norme institutionnelle, s'alignant plus étroitement avec des logiques de ressources informationnelles ou de médias sociaux. Ce sous-groupe de médecins qui prennent leur indépendance grâce aux médias

numériques semble une piste de recherche sous exploitée et bien plus intéressante que celle des alternatives. En somme, cette discussion met en lumière une série de tensions et de contradictions qui caractérisent le champ de l'information de santé. Elle révèle un paysage complexe où les dynamiques de pouvoir, de légitimation et de résistance coexistent et interagissent de manière souvent imprévisible.

### **III.2 - Opérationnalisation et formalisation du concept de “prise de position” sur le web**

La forte corrélation observée entre les variables `emmitter\_type` (type d'émetteur), `media\_type` (type de média) et `themes` (thèmes abordés) a rendu impérative l'élaboration d'une variable composite, que nous avons nommée "position stratégique". Cette variable vise à encapsuler et à synthétiser les dimensions multiples qui caractérisent les prises de position dans le corpus. Elle permet ainsi de transcender les limites des variables individuelles et d'offrir une représentation plus holistique des dynamiques discursives. Pour construire cette variable, nous avons recouru à une Classification Ascendante Hiérarchique (CAH) renforcée par la méthode des K-Means. Cette combinaison méthodologique a été choisie pour sa robustesse et sa capacité à gérer des données de grande dimension. Cette étape a permis de dégager quatre axes stratégiques basés sur 7 clusters distincts (numéroté de 1 à 7) au sein du corpus, chacune représentant une combinaison unique des trois variables initiales. L'analyse des résidus de Pearson a été effectuée à partir d'une série de tables de Khi2, mettant en relation la variable "position stratégique" avec les trois variables constitutives (`emmitter\_type`, `media\_type`, `themes`). Cette démarche statistique a été entreprise pour évaluer l'indépendance conditionnelle entre ces variables et pour identifier des schémas de sur ou sous-représentation. L'analyse des résidus a permis de caractériser la nature spécifique de ces positions stratégiques. En d'autres termes, elle a révélé les qualités intrinsèques qui distinguent chaque position stratégique, en termes de type d'émetteur, de média utilisé et de thèmes abordés. Cette caractérisation a été cruciale pour comprendre les dynamiques sous-jacentes qui régissent les prises de position dans le discours. En somme, la construction de la variable "position stratégique" et l'analyse subséquente des résidus de Pearson ont non seulement enrichi notre compréhension des dynamiques discursives mais ont également permis une opérationnalisation plus nuancée et robuste du concept de prise de positions. Ce faisant, elles ont ouvert la voie à des interprétations plus sophistiquées et à des analyses plus profondes des structures de pouvoir et des stratégies de communication dans le domaine de l'information de santé.

L'examen des positions stratégiques [7] et [4] révèle des dynamiques complexes dans la communication des professionnels de santé. La première position, qui occupe 21,3% du corpus, est fortement institutionnalisée, tandis que la troisième position, représentant 13,7% du corpus, est davantage axée sur le relais d'information par des associations de praticiens. Ensemble, ces deux

positions représentent 35% du corpus total, ce qui souligne leur importance dans le discours sur l'information de santé. Les deux positions convergent sur l'importance de l'éducation thérapeutique, mais diffèrent dans leur approche. La position institutionnelle [7] est plus formelle et autorisée, avec un résidu de Pearson de 4,12 pour l'éducation thérapeutique. En revanche, la position [4] adopte une approche plus pragmatique, axée sur la vie quotidienne avec l'asthme (résidu de 3,92) et la prévention des crises graves (résidu de 2,42). La position institutionnelle [7] privilégie les sites institutionnels (résidu de 21), ce qui est en accord avec son caractère formel et autorisé. La position [4], en revanche, n'indique pas une préférence marquée pour un type de média spécifique, ce qui suggère une plus grande flexibilité dans le choix des plateformes de communication. La surreprésentation des institutions de santé (résidu de 16,44) et des entités publiques (résidu de 8,56) dans la position [7] contraste avec la prédominance des associations de praticiens (résidu de 6,64) dans la position [4]. Cette divergence met en lumière la pluralité des voix au sein des professionnels de santé et souligne la nécessité d'examiner ces positions non pas comme monolithiques, mais comme des entités complexes et nuancées. La juxtaposition de ces deux positions stratégiques offre une fenêtre d'observation fascinante sur la diversité des stratégies de communication adoptées par les professionnels de santé. La position institutionnelle [7] semble être le bastion de l'autorité médicale traditionnelle, où l'information est soigneusement filtrée et diffusée à travers des canaux officiels. En contraste, la position [4] représente une forme de médiation de l'information de santé, où les associations de praticiens jouent un rôle de relais mais aussi d'innovation communicationnel entre les institutions médicales et le besoin du public. La position institutionnelle fournit une base solide et fiable d'informations médicales, tandis que la position de relais enrichit ce discours en y intégrant des perspectives plus pragmatiques et orientées vers le patient. Les professionnels de santé semblent trouver dans la communication numérique une façon d'élargir leurs relations aux patients avec des formes de communication que leur institution délaisse.

La presse, en tant qu'acteur majeur dans le domaine de la communication de santé, se manifeste à travers des positions stratégiques distinctes, qui ensemble représentent près de 29% du corpus total (second territoire en volume). Ces positions révèlent une complexité et une hétérogénéité considérables dans les approches journalistiques, contredisant ainsi l'idée d'un bloc médiatique monolithique. Si les deux sont principalement des acteurs éditoriaux majeurs (avec un résidu de Pearson d'environ 18 pour les médias éditoriaux), deux clusters se dégagent : d'une part la presse spécialisée et médicale et d'autre part, la presse généraliste (hebdomadaire, quotidienne ou régionale). La première représente 18,6% du corpus, est caractérisée par une forte présence de la presse médicale et de la presse spécialisée, notamment les magazines parentaux et féminins. Les résidus de Pearson pour ces deux types de presse sont respectivement de 13,14 et 12,9, indiquant une surreprésentation significative. Les thèmes abordés dans cette position sont souvent à forte charge émotionnelle, notamment les facteurs de crises graves (résidu de 3,04) et les découvertes sensationnelles (résidu de

8,05). Cette position semble s'engager dans le processus d'éducation thérapeutique, mais le fait à travers le prisme de la nouveauté et de l'exceptionnalité propre aux missions d'information des journalistes (après tout news veut dire nouveau). La presse généraliste, qui occupe 10,3% du corpus, domine un autre cluster (résidu de Pearson de 30,72). Cette position est presque entièrement éditoriale et est clairement orientée vers la recherche du scandale et de l'accident, avec un résidu de Pearson de ce type de contenu de 11,53. Ces deux positions stratégiques illustrent les schismes et les nuances qui traversent le paysage médiatique en matière de communication de santé. D'une part, la presse médicale et spécialisée adopte une approche plus nuancée, axée sur l'éducation thérapeutique mais filtrée à travers le prisme de la nouveauté et de l'exceptionnalité. D'autre part, la presse généraliste semble être mue par une logique commerciale, où le sensationnalisme et le scandale sont privilégiés au détriment de l'exactitude et de la nuance. Ces positions stratégiques révèlent une schize sous-jacente dans le rôle de la presse en tant qu'acteur clé dans la communication de santé. Elles soulignent la nécessité d'une analyse plus approfondie pour comprendre les motivations, les biais et les implications éthiques de ces différentes approches journalistiques.

La parole libre et individuelle qui forment la véritable nouveauté du web, notamment celles des patients et des praticiens dans le champ de l'information de santé, jouent un rôle non négligeable dans le paysage de la communication de santé, représentant ensemble 23% du corpus. Ces acteurs manifestent une même position stratégique éditoriale en investissant le web 2.0. On voit se développer une stratégie de l'engagement qui va des réseaux sociaux au blog ou site personnel en passant par les forums spécialisés en santé. Cette position peut être qualifiée de "structurée et engagée", car elle témoigne d'une volonté de participer activement à la communication de santé, tout en adoptant une approche moins institutionnalisée. Les patients et les praticiens cherchent à échanger des conseils pratiques et des expériences personnelles, souvent dans un contexte moins formel et plus communautaire. On distingue tout de même une qualité rédactionnelle plus ou moins aboutie en fonction des types de médias dégageant une certaine hiérarchie au sein de ces prises de paroles quel que soit le média 2.0 utilisé. On distingue une stratégie de création de contenu original, souvent à travers des blogs et des réseaux sociaux, d'une stratégie plus concentrée sur le partage d'informations et l'interaction au sein de communautés en ligne, principalement via des forums. Ce cluster est celui de la sociologie de la communication en matière d'information de santé que nous pouvons désormais quantifier et positionner structurellement dans le champ de l'information de santé.

Le protocole d'analyse de corpus nous a permis de discriminer trois axes majeurs en information de santé à travers 7 clusters liés tant à la nature de l'acteur, aux modes communicationnels et aux discours tenus. Si la parole des patients ont été le champ privilégié de la recherche en SHS et que la parole des professionnels de santé 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 complètement échappé à l'agenda de la recherche. Notre méthode globale, en minimisant les aprioris 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. Parallèlement l'offre des entreprises en produits et service sur le marché de la santé et leur induction en matière d'information de santé est l'autre angle mort des études menées jusque là.

### **III.3 - Mesurer l'influence des acteurs : de la résonance sur les réseaux sociaux à l'authority sur le web**

La question de l'influence des acteurs dans le domaine de la communication de santé sur Internet (cad dire étymologiquement, la capacité à faire bouger le champ social) est d'une importance cruciale, notamment dans le contexte actuel où les informations circulent à une vitesse vertigineuse. Cette influence peut être mesurée selon deux axes principaux : la résonance sociale (audience) multiplié par l'authority sur le web. Chacun de ces axes présente des défis méthodologiques et épistémologiques spécifiques. La résonance sociale, mesurée par le nombre de likes, de partages et de commentaires sur les réseaux sociaux, est un indicateur de l'impact et de la portée des messages. Cependant, plusieurs limitations doivent être prises en compte. Tout d'abord, l'accès à ces données est de plus en plus restreint, notamment en raison de la fermeture des API des réseaux sociaux au cours des 18 mois de notre projet. Cette restriction pose des questions éthiques et méthodologiques, notamment en ce qui concerne la dépendance des chercheurs à l'égard des données propriétaires (Proulx et Rueff, 2018; Cormerais et Lakel, 2019). De plus, notre corpus a révélé que la résonance sociale est un phénomène rare, suivant une loi de distribution exponentielle décroissante (ou loi de Puissance). Cette observation remet en question l'idée d'une viralité omniprésente et suggère que seules quelques unités d'information atteignent un large public. Pour mesurer l'authority web, nous avons adopté une approche basée sur l'analyse structurale des réseaux (Wasserman et Faust, 1994; Scott, 2012). Cette méthode prend en compte les citations explicites dans le corps du texte comme indicateur de légitimité. Cependant, notre étude a révélé une faible connectivité entre les pages web, avec seulement 4,7% des pages étant connectées. Ce constat remet en question l'idée d'un web hyperconnecté et souligne la nécessité de méthodes d'analyse plus nuancées. En combinant ces deux axes avec la prise de position des acteurs, nous pouvons développer une analyse tridimensionnelle de l'influence sur Internet. Cette approche permet non seulement de cartographier les acteurs et leurs stratégies, mais aussi d'étudier les "lieux" d'influence du savoir sur Internet. Elle offre ainsi un cadre d'analyse robuste pour une sociologie numérique des controverses, intégrant à la fois la résonance sociale, l'authority web et les prises de position. La mesure de l'influence sur Internet est un enjeu

complexe qui nécessite une méthodologie rigoureuse et adaptative. En prenant en compte à la fois la résonance sociale et l'autorité web, cette recherche aspire à contribuer de manière significative à la compréhension des dynamiques d'influence dans le domaine de la communication de santé sur Internet.

L'analyse des statistiques de base des interactions sur Facebook et LinkedIn révèle une distribution hautement asymétrique, caractéristique des phénomènes rares (Deheuvels, 1999). Les médianes à zéro pour les likes et les partages sur ces plateformes indiquent que la majorité des publications n'atteignent pas une large audience. Cette observation nous a contraints à recoder les variables en classes qualitatives, subdivisées en trois catégories : absence (0), activité moyenne (1) et exceptionnalité (2). Les tests du chi-carré ( $X^2$ ) montrent une corrélation significative entre les likes et les partages sur Facebook, ainsi qu'entre les partages sur Facebook et LinkedIn. Cette observation suggère que les mécanismes de résonance sociale sont similaires sur les réseaux sociaux grand public et professionnels. De plus, le fait de "liker" et de "partager" semble être deux modalités d'un même phénomène de résonance. L'application de la régression logistique binomiale et du test du  $X^2$  a permis d'identifier les stratégies d'acteurs ayant le plus d'impact sur la résonance sociale. Il en ressort que les stratégies sensationnalistes de la presse quotidienne régionale (PQR) et de la presse spécialisée sont fortement corrélées au succès sur les réseaux sociaux. À l'inverse, les institutions de santé, les associations de praticiens, les blogueurs de santé et les forums de santé sont sous-représentés dans les catégories de grand succès sur Facebook. L'analyse détaillée des facteurs révèle que les blogueurs alternativistes sont les seuls, à part la presse généraliste, à avoir un impact significatif sur la résonance sociale. Les forums et les réseaux sociaux, en revanche, ne semblent pas favoriser les likes sur Facebook, probablement parce que les conversations en ligne ne sont pas propices au partage. Les thèmes sensationnels, tels que les scandales ou les morts accidentelles à l'hôpital, génèrent significativement plus de likes et de partages, conformément au théorème de Fisher qui postule une quantité d'information inversement proportionnelle à la probabilité de son apparition. La résonance sociale, bien que rare, est fortement influencée par des facteurs interdépendants, notamment le type d'acteurs, le média utilisé, et les thèmes abordés. Les stratégies sensationnalistes semblent être les plus efficaces pour générer de la résonance sociale, tandis que les discours médicaux d'éducation thérapeutique et de conseils pour mieux vivre avec des conditions de santé spécifiques, comme l'asthme, sont largement moins percutants à l'échelle du grand public. Ces types de contenus ne bénéficient pas de la viralité des réseaux sociaux grand pourvoyeur d'audience. Cette analyse souligne la complexité de la résonance sociale et la nécessité de hiérarchiser les études. Si les réseaux sociaux sont un indicateurs d'une résonance globale auprès d'un public large, une étude comparée par type de position stratégique d'une audience apportera une meilleurs compréhension par acteurs.

A l'instar de la résonance sociale, le réseau de sites web affiche une très faible interconnexion

avec la même loi de puissance, nécessitant un recodage des paramètres d'analyse des réseaux sociaux en variables binaires. Cette démarche vise à minimiser l'impact des valeurs nulles dans les analyses de régression. L'analyse du X<sup>2</sup> devient alors l'outil statistique le plus approprié pour cette étude. L'analyse du X<sup>2</sup> révèle une corrélation significative entre la nature des acteurs et la probabilité d'être cité. Les praticiens individuels, les associations de patients et la presse médicale sont surreprésentés dans les citations. Étonnamment, aucun thème spécifique ne semble générer plus de citations que d'autres. L'étude a également cherché à comprendre comment différents types de savoirs et modes d'institutionnalisation pourraient influencer l'autorité et la résonance sociale. Les résultats sont sans équivoque : il existe une endogamie entre acteurs partageant des sources de légitimité similaires, ce qui renforce leur isolement. Pour obtenir une variabilité suffisante, l'analyse a été étendue aux domaines d'expression. Les indicateurs de légitimité tels que l'autorité, la centralité et le pagerank ont été examinés. La presse médicale et, dans une moindre mesure, la presse généraliste dominant en termes d'autorité. Les associations de praticiens et les institutions de santé se situent au cœur du réseau de citations, tandis que les individus et les entreprises sont largement sous-représentés. Les acteurs les plus cités sont ceux qui combinent une expertise médicale légitime avec une forme d'intermédiation accessible. La presse médicale et les associations de praticiens en santé jouissent d'une légitimité puissante en termes de co-citation. Le rôle des assurances et mutuelles qui, bien qu'elles renvoient vers des acteurs de santé légitimes, ne bénéficient pas de la même légitimité en raison de leur nature commerciale. Les acteurs qui réussissent à se positionner comme des médiateurs tout en étant légitime dans leur expertise jouent un rôle crucial dans l'établissement de l'autorité web. Cette étude souligne l'importance de prendre en compte à la fois la nature des acteurs, les types de savoirs mobilisés et les modes de communication pour comprendre les dynamiques d'influence et d'autorité dans ce domaine.

L'application de l'algorithme de ModularityClass, bien que largement reconnu pour son efficacité dans l'analyse structurale des réseaux, s'est avérée insuffisante pour dégager des tendances significatives en matière de constitution de communautés endogamiques<sup>4</sup>. Cette insuffisance est en partie due à la faible densité des liens intercommunautaires par rapport aux liens intracommunautaires dans notre corpus. Seuls 23% des 846 sites analysés faisaient partie d'un réseau de citations comprenant au moins trois nœuds, ce qui a rendu l'analyse du X<sup>2</sup> et des résidus de Pearson inopérante pour comprendre les dynamiques endogamiques. Cette faible interconnexion est exacerbée par les critères de sélection de notre outil d'extraction, My Web Intelligence, qui ne retient que les liens qualifiés "d'explicites". Contrairement à l'idée répandue d'un Web 2.0 hautement interconnecté, notre étude révèle une réalité beaucoup plus fragmentée. La majorité des contenus web semblent fonctionner comme des monologues isolés plutôt que comme des éléments d'une toile conversationnelle. Cette constatation remet en question non seulement les postulats sous-jacents de l'analyse des réseaux

---

<sup>4</sup> Pour une application récente voir (Lakel et Le Deuff, 2016)

sociaux mais aussi les méthodologies employées dans l'étude des phénomènes de communication en ligne de beaucoup de recherches en sociologie du web. Elle souligne la nécessité d'adopter des approches méthodologiques plus nuancées et peut-être des outils algorithmiques plus sophistiqués pour capturer la complexité des interactions en ligne. L'échec de cette tentative d'analyse souligne les défis inhérents à la compréhension des structures complexes et souvent contre-intuitives du web. Il met en lumière l'importance de repenser les paradigmes et les outils analytiques pour mieux saisir les dynamiques subtiles qui régissent les interactions en ligne, notamment dans le domaine de la communication de santé. Ce constat d'échec, loin d'être une impasse, ouvre plutôt de nouvelles voies de recherche et d'interrogation qui pourraient s'avérer fructueuses pour les études futures<sup>5</sup>.

---

<sup>5</sup> Cette assertion doit être nuancée par la taille de notre corpus, l'intérêt de travailler dans de plus grands corpus devrait permettre d'étudier une partie plus importante de pages web interconnectées.

# Conclusion générale

Le projet de recherche s'engage dans un examen des intersections complexes entre les paradigmes médicaux traditionnels et les paradigmes informationnels à l'ère numérique, tout en intégrant des perspectives interdisciplinaires issues de la santé publique, de la sociologie de la communication, et des études numériques. Le projet de recherche a voulu partir d'une critique méthodologique rigoureuse des approches conventionnelles pour évaluer la qualité de l'information de santé sur Internet. Il a mis en question la fiabilité et la pertinence de ces méthodes en exposant leurs limites, notamment leur incapacité à prendre en compte les dynamiques à la fois structurales et changeantes du pouvoir et de l'autorité dans le contexte numérique. Ce projet a tenté de dépasser les limites méthodologiques en introduisant de nouveaux cadres socio-épistémologiques pour comprendre la manière dont l'information de santé est construite, partagée, et utilisée en ligne. Ces cadres permettent une analyse plus nuancée et multidimensionnelle des interactions entre les différents acteurs dans l'espace numérique, y compris les patients, les professionnels de la santé, et d'autres intermédiaires de l'information. Le projet utilise My Web Intelligence comme outil d'analyse numérique pour explorer ces dynamiques, offrant ainsi une approche plus granulaire et contextuelle pour étudier les controverses et les discours en matière de santé en ligne.

En effet l'innovation en matière de digital methods est le second axe majeur de cette recherche portant une attention particulière aux dynamiques de pouvoir en mutation entre les patients et les professionnels de la santé dans un environnement numérique. En utilisant des méthodes quantitatives sur des métadonnées numériques complexes, le projet met en lumière la diversité des prises de positions stratégiques des actants. Il montre que ces dynamiques sont loin d'être unidimensionnelles; elles sont plutôt influencées par une multitude de facteurs, notamment les plateformes utilisées, le type de contenu partagé, et les structures réseau dans lesquelles ces acteurs sont intégrés. Par conséquent, il révèle que les patients ne sont pas de simples récepteurs passifs d'information mais des acteurs qui se positionnent et participent à la construction et à la diffusion de l'information de santé en ligne. En somme, ce projet de recherche offre une contribution à la littérature sur l'information de santé en ligne en proposant des méthodologies et des cadres théoriques innovants. La recherche a voulu élargir notre compréhension des dynamiques complexes qui régissent le partage et l'utilisation de l'information de santé sur Internet, et offre des pistes pour des interventions plus efficaces et des politiques plus éclairées dans ce domaine.

L'étude de cas se concentre spécifiquement sur les conditions de diabète et d'asthme chez l'enfant et apporte des contributions à la compréhension de l'information de santé perçue par un patient qui userait de Google pour s'informer (plus des 2/3 des patients). En utilisant des méthodes numériques

avancées, l'étude opérationnalise et quantifie le concept de "prise de position," permettant ainsi une évaluation plus rigoureuse des types de discours et des acteurs présents dans ce domaine spécifique de la santé en ligne. Mais surtout de leur influence. L'étude identifie plusieurs types de locuteurs actifs dans l'information de santé en ligne, allant des patients aux professionnels de la santé. Elle analyse ensuite comment leur autorité et leur influence varient en fonction de plusieurs facteurs contextuels, tels que la plateforme sur laquelle l'information est partagée, le type de contenu et leur intégration dans des réseaux plus larges. Par exemple, les professionnels de la santé tendent à avoir une influence plus importante dans des environnements spécifiques, tandis que les patients peuvent avoir une résonance plus grande sur des plateformes de médias sociaux. Dans le cas du diabète et de l'asthme chez l'enfant, les résultats montrent que les discours dominants sont souvent produits par des institutions médicales et des professionnels de la santé qui restent maîtres des horloges. Alors les voix des patients et des parents sont relativement moins présentes et moins influentes, la presse médicale, spécialisée (surtout féminine et parentale) et généraliste assure sa fonction d'intermédiation entre le patient et le soignant (bien que cette intermédiation soit biaisée par une recherche permanente du scoop).

La recherche s'inscrit résolument dans une démarche interdisciplinaire qui traverse les frontières de la sociologie numérique, des sciences de l'information, et de la santé publique. Cette combinaison permet une approche plus holistique à la complexité de l'information de santé en ligne, une caractéristique souvent absente dans des études plus traditionnelles qui sont confinées à des domaines spécifiques, tels que la médecine basée sur les preuves ou les études de communication. L'utilisation de My Web Intelligence pour l'analyse numérique place cette étude à la pointe des innovations en digital methods. Le recours à des méthodes numériques permet d'aborder des questions souvent négligées, comme les dynamiques de pouvoir et les structures d'interaction, en fournissant des mesures quantifiables et des perspectives granulaires. L'étude de cas sur le diabète et l'asthme chez l'enfant ouvre la perspectives de contributions empiriques dans un domaine qui a été principalement exploré à travers des lentilles restreintes. Contrairement aux études antérieures qui se concentrent souvent sur l'information de santé en tant que catégorie monolithique, cette recherche cible des conditions médicales spécifiques. En mettant en lumière un spectre large d'acteurs, elle a permis de dégager une vision structurale et holistique du champ, favorisant par la suite des plongées dans des sous-espaces spécialisés pour dégager des dynamiques spécifiques qui sont souvent masquées dans des analyses plus globales.

En résumé, cette recherche se positionne dans le paysage international de la recherche en information de santé en ligne. Elle enrichit le corpus de connaissances en introduisant des méthodologies et des perspectives interdisciplinaires, tout en fournissant les moyens de construire des données empiriques spécifiques. Ce faisant, elle établit des liens entre des domaines de recherche souvent traités de manière isolée et propose de nouvelles voies pour la recherche en SHS assisté par

ordinateur. La perspective de l'auteur est de poser les fondations d'un programme pour une série de recherches futures qui pourraient s'orienter selon plusieurs axes. Tout d'abord, elle met en évidence la nécessité de constituer des corpus de données plus représentatifs, notamment en ce qui concerne l'information de santé en ligne. Cette préoccupation rejoint des défis méthodologiques majeurs dans les sciences humaines et sociales, notamment les questions autour de l'accès, de l'interopérabilité, et de la propriété des données. Ce problème est d'autant plus pertinent dans le contexte de l'essor des humanités numériques, où la manipulation de grandes quantités de données devient une pratique courante mais complexe en termes d'accès. Deuxièmement, cette recherche soulève des questions importantes sur la politique publique de la donnée en sciences humaines et sociales. Elle pointe vers des défis relatifs aux limites techniques, économiques, et symboliques de l'accès à des données pertinentes. Ces défis appellent à une réflexion plus poussée sur la gouvernance des données, qui pourrait constituer un sujet de recherche en soi, notamment dans le cadre des politiques de santé publique et de la réglementation de l'information en ligne. Troisièmement, l'étude indique un besoin urgent d'explorer plus en profondeur l'information de santé telle qu'elle est médiatisée par les réseaux sociaux. Alors que cette étude s'est focalisée sur les résultats des moteurs de recherche, elle met en lumière l'importance des réseaux sociaux dans la consommation de l'information. Ce constat appelle à des études complémentaires qui pourraient utiliser des méthodologies différentes pour examiner l'information de santé sur ces plateformes, telles que l'analyse de réseaux sociaux ou les études ethnographiques en ligne. Quatrièmement, l'étude souligne la nécessité d'études comparatives pour consolider les résultats obtenus. Le recours à des méthodologies différentes ou à des échantillons plus larges pourrait permettre de valider, de nuancer, ou de compléter les conclusions tirées de cette recherche.

En somme, cette étude sert de point de départ pour une variété de recherches futures qui pourraient aborder les questions complexes et interdisciplinaires posées par l'information de santé en ligne. Elle invite à une réflexion plus globale sur les méthodologies, les politiques de données et les dynamiques sociales et épistémologiques qui influencent ce domaine.

## Bibliographie

- Akrich, M., & Méadel, 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., & Méadel, C. (2007). De l'interaction à l'engagement : les collectifs électroniques, nouveaux militants de la santé, From Interaction to Activism: E-Health Communities. *Hermès, La Revue*, (47), 145-153.
- Akrich, M., Méadel, C., Rémy, C., & Vergnaud, F. (2009). *Les patients et l'information : le cancer au risque d'internet*. 217.
- 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>
- Banti, T., Carsin, A., Chabrol, B., & Fabre, A. (2016). Diversification alimentaire : l'information disponible sur Internet est-elle en accord avec les recommandations des sociétés savantes ? *Archives de Pédiatrie*, 23(7), 706-713. <https://doi.org/10.1016/j.arcped.2016.04.017>
- Banti, T., Fievet, L., & Fabre, A. (2017). Gluten : l'information disponible sur Internet est-elle en accord avec les recommandations des sociétés savantes ? *Archives de Pédiatrie*, 24(10), 977-985. <https://doi.org/10.1016/j.arcped.2017.08.007>
- Barats, C. (2017). *Manuel d'analyse du web - 2e éd.* - NP. Armand Colin.
- Bastard I., Cardon D., Fouetillou, G., Prieur, C. et Raux, S. (2013). « Les sciences sociales et les données indiscretes du web », *Le Monde science et techno*. En ligne sur : <[http://www.lemonde.fr/sciences/article/2013/12/16/les-sciences-sociales-et-les-donnees-indiscretes-du-web\\_4335257\\_1650684.html](http://www.lemonde.fr/sciences/article/2013/12/16/les-sciences-sociales-et-les-donnees-indiscretes-du-web_4335257_1650684.html)>, consulté le 23 mai 2014.
- Bastard, I., Cardon, D., Charbey, R., Cointet, J.-P., & Prieur, C. (2017). « Facebook, Pour Quoi Faire ? » dans *Sociologie*, N° 1, 8. En ligne sur <<http://sociologie.revue.org/3040>>
- 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.
- Baya-Laffite, N., & Cointet, J.-P. (2014). Cartographier la trajectoire de l'adaptation dans l'espace des négociations sur le climat, *Réseaux*, (188), 159 -198. <https://doi.org/10.3917/res.188.0159>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Boullier, D. (2016). *Sociologie du numérique*. Paris: Armand Colin.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2), 163-177. <https://doi.org/10.1080/0022250X.2001.9990249>
- 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., 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, Der Arzt, sein Patient und seine Kollegen. Eine neue Näherungsweise der therapeutischen Beziehung. *Revue française de sociologie*, 46(3), 443-467. <https://doi.org/10.3917/rfs.463.0443>
- Chaudiron, S., & Ihadjadene, M. (2010). De la recherche de l'information aux pratiques informationnelles. *Études de communication. langages, information, médiations*, (35), 13-30. <https://doi.org/10.4000/edc.2257>
- Convert, B., & Demailly, L. (2003). Internet et les professions de sante. *Réseaux*, no 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 », Revue Etudes Digitales 6, Garnier classiques.
- Cormerais, Franck. Le Deuff, Olivier (2014) « Le lettré du numérique dans la reconfiguration du savoir contemporain », *Digital Intelligence 2014*, Nantes, 2014. En ligne sur : <[www.univ-nantes.fr/servlet/com.univ.collaboratif.util.LectureFichier?CODE\\_FICHER=1415428057047&ID\\_FICHE=717841](http://www.univ-nantes.fr/servlet/com.univ.collaboratif.util.LectureFichier?CODE_FICHER=1415428057047&ID_FICHE=717841)>, consulté le 02 juin 2017.
- Espitalier, F., Barbaz, M., Fusciardi, J., & Laffon, M. (2013). Péridurale obstétricale : qualité de l'information médicale en français sur Internet en 2009 et 2010. *Annales Françaises d'Anesthésie et de Réanimation*, 32(9), 554-559. <https://doi.org/10.1016/j.annfar.2013.05.003>
- Espitalier, F., De Lamer, S., Barbaz, M., Remérand, F., Fusciardi, J., & Laffon, M. (2014). Qualité de l'information médicale et péridurale obstétricale : quoi de neuf en 2013 ? *Annales Françaises d'Anesthésie et de Réanimation*, 33, Supplement 2, A351-A352. <https://doi.org/10.1016/j.annfar.2014.07.647>
- Evrard, A.-S., Guertin, L., Remacle, M., Jamart, J., & Lévêque, M. (2009). Information Internet en langue française en oncologie ORL. *Annales d'Otolaryngologie et de Chirurgie Cervico-faciale*, 126(3), 99-111. <https://doi.org/10.1016/j.aorl.2009.05.001>
- Ferraris, Maurizio (2012). *Documentality: Why It Is Necessary to Leave Traces*. Cambridge : Oxford University Press
- Flichy, P. (1995). *L'innovation technique : Récents développements en sciences sociales, vers une nouvelle théorie de l'innovation*. Paris: La Découverte.
- Frémont, P., Labrecque, M., Légaré, F., Baillargeon, L., & Misson, L. (2001). Evaluation des sites web médicaux. Fidélité interobservateur et intraobservateur d'un outil d'évaluation. *Canadian Family Physician*, 47, 2270-2278.
- Ghiglione, R. (1986). *L'homme communiquant*. Paris, France: A. Colin, DL 1986.

- Hargrave, D., Bartels, U., Lau, L., Esquembre, C., & Bouffet, É. (2003). Évaluation de la qualité de l'information médicale francophone accessible au public sur internet : application aux tumeurs cérébrales de l'enfant. *Bulletin du Cancer*, 90(7), 650-655.
- Jendoubi, S. (2016). *Influencers characterization in a social network for viral marketing perspectives* (Phdthesis, Université Rennes 1). Consulté à l'adresse <https://tel.archives-ouvertes.fr/tel-01665815/document>
- Kleinberg, J. M. (1999). Authoritative Sources in a Hyperlinked Environment. *J. ACM*, 46(5), 604–632. <https://doi.org/10.1145/324133.324140>
- Koszowska-Nowakowska, P., & Renucci, F. (2012). L'hypertexte : approches expérimentale et herméneutique. *Les Cahiers du numérique*, Vol. 7 (3), 71-91.
- Lafon, B., & Pailliar, I. (2007). Malades et maladies dans l'espace public. *Questions de communication*, (11), 7-15. <https://doi.org/10.4000/questionsdecommunication.7319>
- Lakel Amar, « Interconnexion et identités électroniques : vers un système de fichiers répartis » in Terminal n° 99-100, dossier spécial Administration électronique : où en sommes-nous ?, 2007
- Lakel, A., & Massit-Folléa, F. (2007). Société civile et gouvernance de l'internet : la construction d'une légitimité ambiguë, Civil Society and Internet Governance: Building an Ambiguous Legitimacy. *Hermès, La Revue*, (47), 167-176.
- 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., & Deuff, O. L. (2017). À quoi peut bien servir l'analyse du web ? *Les Cahiers du numérique*, 13 (3), 39-62.
- Latour, B. (2007). *Reassembling the Social: An Introduction to Actor-Network-Theory* (1 edition). Oxford: Oxford University Press.
- Latour, B. (2010, février 19). *Networks, Societies, Spheres: Reflections of an Actor-Network Theorist*. Présenté à International Seminar On Network Theory: Network Multidimensionality In The Digital Age. Consulté à l'adresse <https://hal-sciencespo.archives-ouvertes.fr/hal-00972865/document>
- LATOUR, B. (2014). *Changer de société, refaire de la sociologie*. La Découverte.
- Legros, M. (2010). Étude exploratoire sur les blogs personnels santé et maladie. *Santé Publique*, 21(hs2), 41-51.
- Lévêque, M., Dimitriu, C., Gustin, T., Jamart, J., Gilliard, C., & Bojanowski, M. W. (2007). Évaluation de l'information sur Internet destinée aux patients francophones en neuro-oncologie. *Neurochirurgie*, 53(5), 343-355. <https://doi.org/10.1016/j.neuchi.2007.07.033>
- Lévi-Strauss, C. (1990). *La pensée sauvage* (Revised). Paris: Pocket.
- Lipsyc, C., & Ihadjadene, M. (2014). Architecture de l'information et éditorialisation. *Études de communication*, n° 41(2), 103-118.
- Méadel, 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.
- Perriault, J., & Schaeffer, P. (2008). *La logique de l'usage : Essai sur les machines à communiquer*. s.l.: Editions L'Harmattan.
- Poibeau, T. (2014). Le traitement automatique des langues pour les sciences sociales, *Réseaux*, (188), 25-51. <https://doi.org/10.3917/res.188.0025>
- 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)
- 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.
- Rossi, M. (2017). *Graph Mining for Influence Maximization in Social Networks* (Phdthesis, Université Paris-Saclay). Consulté à l'adresse <https://pastel.archives-ouvertes.fr/tel-01688005/document>
- Saemmer, A. (2012). Hypertexte et irradiation iconique. *Les Cahiers du numérique*, Vol. 7 (3), 47-69.
- Saemmer, A. (2015). Hypertexte et narrativité. *Critique*, n° 819-820 (8), 637 -652.
- Scott, J. (2017). *Social Network Analysis* (4 edition). Thousand Oaks, CA: SAGE Publications Ltd.
- 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>
- Vercher, E., & Touboul, A. (2010). *Diffusion de l'information et médiation dans le domaine de la santé : nouvelles pratiques, nouvelles règles*. 9.
- Vitali Rosati, M. (2016). Qu'est-ce que l'éditorialisation? *Sens public*. <https://doi.org/10.7202/1043383ar>
- Vitali-Rosati, Merello, Sinatra, Michael E. (dir.) (2014). *Pratiques de l'édition numérique. Nouvelle édition*. Montréal : Presses de l'Université de Montréal, 2014. Disponible sur Internet : <<http://books.openedition.org/pum/306>>.
- Vivien, A., Kowalski, V., Chatellier, A., Babin, E., Bénateau, H., & Veyssière, A. (2017). Qualité de l'information des sites francophones destinés au public pour le dépistage des cancers oraux. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 118(1), 20-28. <https://doi.org/10.1016/j.jormas.2016.10.003>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (1 edition). Cambridge ; New York: Cambridge University Press.