

# Actualité des méthodes de recherche en sciences sociales sur les pratiques informationnelles

---

**Serge Proulx**

**Julien Rueff**



CENTRE D'ÉTUDES SUR LES MÉDIAS



# **Actualité des méthodes de recherche en sciences sociales sur les pratiques informationnelles**

**Serge Proulx**

**Julien Rueff**

Révision: Marie-Hélène Lavoie  
Infographie: Diane Trottier  
Coordination: Sébastien Charlton  
ISBN 978-2-922008-60-9

Novembre 2018

Centre d'études sur les médias  
Pavillon Casault (5604)  
Université Laval  
Québec (Québec) G1V 0A6  
Téléphone: 418-656-3235  
Télécopieur: 418 656-7807  
Adresse électronique: CEM@com.ulaval.ca  
Site Internet: <http://www.cem.ulaval.ca>

Directrice du Centre: Colette Brin

Droits d'auteur et droits de reproduction : Toutes les demandes de reproduction doivent être acheminées à Copibec : 514-288-1664 ou 1 800 717-2022, [licences@copibec.qc.ca](mailto:licences@copibec.qc.ca).

# Résumé synthétique

**P**rofondément affecté par le numérique, notre paysage médiatique n'en finit pas de dérouter autant les journalistes professionnels, les producteurs d'informations que les propriétaires de médias traditionnels. Les chercheurs en sciences sociales désireux de donner une intelligibilité aux transformations actuelles des pratiques informationnelles n'échappent pas à cette tendance. Force est de constater une diversification des sources d'information, une fragmentation des pratiques de réception, une multiplication des supports matériels (téléphones intelligents, tablettes, ordinateurs et autres objets connectés), mais aussi une généralisation de la pluriactivité (*multitasking*) dans les manières de s'informer. Ces évolutions concourent à produire un environnement dans lequel les pratiques informationnelles (manières de s'informer) se montrent difficilement saisissables, en raison de leur instabilité grandissante. Ce rapport dresse un portrait des différentes stratégies méthodologiques à partir desquelles les chercheurs en sciences sociales peuvent aujourd'hui espérer construire des connaissances sur ces manières de s'informer. Cette analyse mène ultimement à la formulation de six recommandations au Centre d'études sur les médias (CEM).

## **A. Défis méthodologiques et épistémologiques**

Ce rapport met en lumière les problèmes et les enjeux des aspects méthodologiques dans l'étude des pratiques informationnelles contemporaines. Pour ce faire, il s'appuie sur l'expertise de 24 chercheurs francophones en sociologie, en communication et en gestion. Puisqu'ils ont tous travaillé à la conception de dispositifs méthodologiques pour collecter et analyser des données sur des activités en ligne, leur parole s'est avérée précieuse pour nous aider à saisir les principaux enjeux des débats scientifiques contemporains autour des méthodes de

recherche sur les manières de s'informer. Trois dimensions ont été abordées de façon récurrente par les personnes que nous avons interrogées : le diagnostic d'une complexification des pratiques informationnelles aujourd'hui, l'expérimentation méthodologique et, enfin, l'existence de problèmes épistémologiques.

La première dimension a trait à l'opacité grandissante des phénomènes empiriques liés à l'information d'actualité en ligne. Ce contexte incite les chercheurs – et c'est la deuxième dimension ressortant avec évidence de nos entretiens – à s'engager dans un « bricolage méthodologique », cette expression renvoyant non pas à des errements approximatifs, mais à des pratiques faisant appel à l'ingéniosité et à l'inventivité. La complexification des manières de s'informer invite non seulement à une réévaluation rigoureuse des techniques usuelles de collecte et d'analyse de données en sciences sociales, mais aussi à une élaboration de nouvelles stratégies méthodologiques. Les experts de notre panel ont, à ce propos, plaidé avec insistance en faveur d'une combinaison, voire d'une hybridation, des approches qualitatives et computationnelles pour appréhender les pratiques informationnelles contemporaines.

La dernière dimension relève de considérations épistémologiques. Au regard de ce que d'aucuns appellent le « tournant computationnel » en sciences sociales, les chercheurs de notre panel ont fréquemment opéré une distinction entre les stratégies méthodologiques permettant, d'un côté, de *mesurer* les manières de s'informer actuelles et, de l'autre, d'en proposer une *compréhension*. Cette dichotomie permet de prendre conscience de l'existence de plusieurs modes d'intelligibilité dans les travaux en sciences sociales, et en particulier, de la nécessité de ne pas substituer un peu trop hâtivement la mesure à la compréhension. L'existence de ce « tournant computationnel » en sciences sociales a été mise en question dans nos entretiens. Cette interrogation a fait apparaître des positions contrastées parmi les chercheurs. Enfin, certains chercheurs interrogés ont évoqué les problèmes posés par une discussion sur les méthodes qui se ferait sans ancrage disciplinaire et sans référence aux problématiques concrètes que les recherches visent à éclairer. L'enjeu devient alors de mettre en question la légitimité d'une autonomisation de ces débats par rapport aux grandes problématiques propres aux différentes sciences sociales.

## B. Quatre idéaux-types de méthode de recherche

Ce rapport construit une typologie des principales méthodologies dans les recherches sur les pratiques informationnelles actuelles. Quatre grandes approches méthodologiques ont été distinguées : les méthodes conventionnelles (qualitatives ou quantitatives), l'ethnographie en ligne, les méthodes computationnelles appliquées aux traces des *big data* et les méthodes numériques (*digital methods*). Si les méthodes conventionnelles – qualitatives et quantitatives – sont largement diffusées dans les formations universitaires en sciences sociales ou dans des manuels et, à ce titre, bien connues, il n'en est pas de même pour les trois autres idéaux-types de méthode en lien avec le numérique. L'ethnographie en ligne – que la chercheuse Christine Hine a contribué à populariser depuis les années 2000 – vise à produire des « descriptions denses et détaillées » de la réalité observée en prenant acte de la prégnance de la médiatisation numérique dans les pratiques sociales. L'ethnographie en ligne s'inscrit dans la continuité de l'ethnographie traditionnelle, mais elle s'efforce d'adapter ses techniques d'enquête au contexte actuel. Cette approche méthodologique se caractérise par ailleurs par sa pluralité, son adaptabilité et sa nature qualitative. Les méthodes computationnelles exploitant les *big data* – ayant fait émerger notamment les approches de type *big methods* pour reprendre l'expression de Noortje Marres – sont nées de la rencontre des puissantes technologies computationnelles et de la disponibilité des traces des *big data*. Ces dernières se caractérisent par leur volume, leur vélocité, leur exhaustivité, la finesse de leur granularité, leur flexibilité. Ce troisième idéal-type de méthode s'accompagne par ailleurs de l'utilisation d'outils de visualisation pour rendre compte des traces extraites des plateformes en ligne. L'analyse des traces des *big data*, dans le cadre de ces méthodes computationnelles, repose notamment sur la recherche de corrélations entre différents items de ces très grands corpus de données. Le quatrième idéal-type renvoie aux méthodes numériques (*digital methods*). Quoique variées, ces dernières ont pour dénominateur commun de vouloir se saisir des opportunités offertes par les objets nativement numériques, par le désir de « suivre le medium » en le considérant en tant que ressource méthodologique, mais à condition de réorienter

(*repurposing*) ces mesures dans le cadre de problématiques sociétales. Nous avons distingué deux grandes familles de méthodes numériques: d'un côté, les *digital methods* telles que définies par la *Digital Methods Initiative* coordonnée par Richard Rogers (Amsterdam); et de l'autre, les méthodes numériques quali-quantitatives. Nous croyons cependant que ces deux manières de faire sont en train de se fusionner.

### C. Étude de cas

Pour chaque idéal-type, nous avons repéré une étude de cas concernant spécifiquement les manières de s'informer au temps du numérique. Nous offrons une présentation analytique de chacune des études. En mettant l'accent sur les potentialités et limites respectives des stratégies méthodologiques mobilisées, nous avons retenu une grille d'analyse à cinq dimensions, à savoir: les présupposés épistémologiques de chaque étude de cas; leurs assises théoriques; leurs techniques de collecte et d'analyse; un aperçu des résultats obtenus; et les limites du dispositif méthodologique. Pour les méthodes conventionnelles, nous avons retenu une étude du Pew Research Center sur les usages des médias numériques (méthodes quantitatives) et une enquête menée par Guillaume Le Saulnier sur la lecture de la presse en ligne (méthodes qualitatives). Les recherches menées par Coralie Le Caroff pour sa thèse de doctorat ont été mobilisées pour exemplifier l'ethnographie en ligne: ce travail portait sur les usages sociopolitiques de l'actualité sur le réseau Facebook. Pour les méthodes computationnelles appliquées aux traces des *big data*, nous avons choisi une recherche réalisée en Australie par Axel Bruns, Tim Highfield et Stephen Harrington qui concernait le partage d'informations d'actualité sur Twitter. Enfin, nous abordons l'adoption des méthodes numériques quali-quantitatives en décrivant une recherche réalisée par Irène Bastard qui concerne le partage de l'information d'actualité sur Facebook.



## D. Recommandations au Centre d'études sur les médias

De ce qui précède, nous aimerions soumettre six recommandations au Centre d'études sur les médias concernant les dispositifs méthodologiques des recherches en sciences sociales sur les pratiques informationnelles actuelles. Ces recommandations demeurent volontairement générales et ne sauraient être associées à des préconisations à respecter strictement. Dans la mesure où les techniques de collecte et d'analyse de données varient nécessairement en fonction de la problématique et du contexte de la recherche, il ne paraît pas souhaitable d'enfermer la réalisation des études sur les manières de s'informer dans des modèles rigides. Insistons sur ce point : il n'existe pas de méthode de recherche « universelle » permettant de traiter toutes les questions scientifiques ou embrassant toutes les dimensions des pratiques informationnelles.

1. Le développement de stratégies méthodologiques hybrides, articulant des techniques d'enquête qualitatives et quantitatives, nous apparaît plus que souhaitable pour permettre la description, la mesure et la compréhension des pratiques informationnelles actuelles.
2. Les méthodes numériques (*digital methods*) constituent une voie prometteuse pour l'étude des manières de s'informer aujourd'hui. Nous préconisons par conséquent une articulation entre les méthodes conventionnelles, que celles-ci soient qualitatives ou quantitatives, et les méthodes computationnelles appliquées aux traces des *big data*.
3. Dans la mesure où les méthodes numériques (*digital methods*) ouvrent des perspectives intéressantes pour la recherche sur l'information médiatique et le journalisme, nous suggérons la formation d'équipes interdisciplinaires, rassemblant des chercheurs en sciences sociales, en informatique et en sciences des données (*data scientists*).
4. En cohérence avec les recommandations précédentes, nous décourageons le recours unique et exclusif à des méthodes computationnelles appliquées aux traces des *big data*. Les *big methods* ne peuvent qu'apporter aux manières de s'informer

contemporaines une intelligibilité limitée, fondée sur la mesure, mais déconnectée des logiques de signification des acteurs ou des contextes spécifiques dans lesquels se produisent leurs comportements.

5. La mobilisation des méthodes numériques (*digital methods*) n'est pas sans difficulté. Certains risques doivent être pris en considération avant l'adoption d'une telle stratégie méthodologique. La volatilité du design des plateformes en ligne constitue un premier problème, dans la mesure où les chercheurs n'ont d'autre choix que de «subir» les évolutions rapides – et parfois radicales – des caractéristiques sociotechniques des médias numériques. De la même manière, les chercheurs doivent avoir conscience des bouleversements que peuvent engendrer des changements sur le plan des modalités d'accès aux données des plateformes en ligne.
6. Sur le plan éthique, nous préconisons la mise en place de dispositifs méthodologiques respectueux de la protection des données personnelles des enquêtés, d'une part, et passant par l'obtention du consentement des individus, d'autre part.

# Table des matières

<b>Résumé synthétique</b> .....	<b>5</b>
A. Défis méthodologiques et épistémologiques .....	5
B. Quatre idéaux-types de méthode de recherche .....	7
C. Étude de cas .....	8
D. Recommandations au Centre d'études sur les médias.....	9
<b>1 Introduction</b> .....	<b>13</b>
<b>2 Analyser les pratiques informationnelles dans un contexte d'instabilité</b> .....	<b>19</b>
1. Une complexification des pratiques informationnelles.....	22
2. Le « bricolage » méthodologique .....	27
2.1 Repenser les axes d'analyse .....	28
2.2. Les phénomènes d'influence dans les processus de circulation de l'information .....	30
2.3 Quelles données employer pour étudier les pratiques informationnelles actuelles? .....	32
2.4 Quelques définitions.....	35
2.5 L'hybridation des méthodes qualitatives et computationnelles .....	39
3. Quelques considérations épistémologiques.....	41
3.1 Quelle intelligibilité peut-on espérer de ces choix méthodologiques? .....	41
3.2 Peut-on parler « d'un tournant computationnel » en sciences sociales? .....	46
3.3 Une discussion méthodologique sans ancrage disciplinaire? .....	49
<b>3 Les méthodes de recherche en sciences sociales en lien avec le numérique</b> .....	<b>51</b>
1. Idéaux-types des stratégies méthodologiques .....	53
1.1 Les méthodes conventionnelles .....	55
1.2 Les ethnographies virtuelles.....	55

1.3	Les méthodes computationnelles exploitant les big data (approche dite « big methods »).....	58
1.4	Les « Digital Methods » de la Digital Methods Initiative .....	62
1.5	Les méthodes numériques quali-quantitatives .....	68
2.	Enjeux organisationnels, économiques, politiques et éthiques.....	71
2.1	Enjeux organisationnels: la collaboration entre chercheurs en sciences sociales et informaticiens .....	71
2.2	Enjeux « politiques » de l'accès aux traces des big data.....	74
2.3	Enjeux économiques: la dépendance des chercheurs au secteur marchand .....	78
2.4	Problèmes éthiques des recherches en lien avec le numérique.....	79
<b>4</b>	<b>Potentialités et limites des méthodes de recherche en lien avec le numérique.....</b>	<b>83</b>
1.	Les méthodes conventionnelles.....	85
1.1	Une étude quantitative des usages des médias sionumériques .....	85
1.2	Une étude qualitative sur la lecture de la presse en ligne .....	89
2.	Une ethnographie en ligne des usages sociopolitiques de Facebook.....	94
3.	Une étude sur le partage d'information par Twitter .....	103
4.	Une étude quali-quantitative sur les usages de Facebook ..	114
<b>5</b>	<b>Conclusions .....</b>	<b>127</b>
	Quels sont les principaux points à retenir de ce rapport?.....	127
	Une analyse comparative des idéaux-types de méthode .....	131
	Recommandations au Centre d'études sur les médias.....	136
	<b>Glossaire .....</b>	<b>139</b>
	<b>Bibliographie .....</b>	<b>143</b>
	<b>Annexe – Liste des personnes interviewées .....</b>	<b>151</b>

# 1

## Introduction

Notre démarche a comme point de départ un constat: les chercheurs en sciences sociales éprouvent aujourd'hui des difficultés méthodologiques particulières lorsqu'ils font enquête auprès des personnes et des publics pour connaître précisément leurs habitudes contemporaines en matière d'information<sup>1</sup>. Avant la pénétration massive d'internet, les méthodes d'enquête conventionnelles pouvaient être facilement utilisées pour tracer un portrait relativement juste des habitudes des personnes en matière de consommation d'information. Ainsi, le Centre d'études sur les médias a réalisé, tous les deux ans depuis 2007, auprès d'un même panel d'environ 500 personnes, des enquêtes documentant les manières de s'informer par la presse imprimée et les magazines, la radio, la télévision, et par internet. Ces enquêtes par questionnaire ont été réalisées au moyen d'une technique de cueillette de données que le CEM a gardée stable le plus possible au fil des ans, afin de faciliter les comparaisons. Les résultats de ces enquêtes longitudinales ont montré notamment que l'adoption par les Québécois de modes d'information en ligne prenait une ampleur de plus en plus grande. C'est dans ce contexte que le besoin de dispositifs méthodologiques plus

---

1. Il s'agit ici de l'information au sens journalistique du terme qui concerne en particulier les « actualités » et les thématiques (sociales, économiques, culturelles, politiques) faisant l'objet de débats et controverses dans la sphère publique.

sophistiqués s'est fait sentir. Alors que les enquêtes par questionnaire se fondent sur les *déclarations* des répondants à propos de leurs manières de s'informer, comment imaginer en effet que ces personnes soient encore capables de retracer consciemment leurs parcours sur internet?

L'environnement informationnel s'est effectivement complètement transformé à l'heure du numérique. Comme nous le verrons dans le chapitre suivant, le paysage médiatique s'est significativement modifié notamment à travers l'abondance et la diversification des sources d'information, la fragmentation exacerbée des pratiques de réception – les récepteurs individuels devenant en même temps des producteurs de messages ou des « passeurs » dans une dynamique de partage –, la multiplication des supports matériels (téléphones intelligents, tablettes, ordinateurs et autres objets connectés), une généralisation de la pluriactivité (*multitasking*) dans les manières de s'informer, etc. Ce contexte de mutation aboutit à l'instabilité extrême des pratiques informationnelles, cette incertitude s'avérant « dynamique » dans la mesure où le rythme d'accélération du changement suit une courbe ascendante. Par ailleurs, en 2016, nous avons enquêté auprès d'un sous-échantillon des répondants du panel ayant déclaré avoir changé leurs habitudes d'information avec l'avènement du numérique. Avec cette enquête, nous avons cherché à ce que ces répondants nous expliquent leurs motivations à modifier leurs habitudes et, également, leurs perceptions de ces changements (Giroux, Proulx, Souissi, 2017).

Plusieurs enquêtes publiées depuis plus d'une décennie – qu'il s'agisse de recherches réalisées au Québec, en Amérique du Nord ou en Europe – confirment que les individus effectuent un passage significatif vers l'environnement numérique dans leurs manières de s'informer. L'intérêt heuristique consiste aujourd'hui à pouvoir documenter de manière fine cette transformation des pratiques informationnelles. Il s'avère pertinent de se demander quelles approches méthodologiques en sciences humaines et sociales (SHS) pourraient être mobilisées pour en apprendre

davantage sur ces pratiques<sup>2</sup>. Ces approches nouvelles (combinées aux anciennes méthodes) pourraient éventuellement produire des résultats d'enquêtes propres à nous éclairer sur les enjeux organisationnels (économiques, culturels, politiques) liés à ce contexte de mutation du paysage médiatique, considérant le fait que la période actuelle (à l'horizon incertain) est vécue comme difficile par plusieurs journalistes professionnels et par certains propriétaires de médias traditionnels producteurs d'information.

Notre démarche de recherche – une « métaenquête » en quelque sorte – a consisté d'abord à interviewer un certain nombre de chercheurs en sciences sociales s'intéressant particulièrement aux pratiques informationnelles liées à l'usage des plateformes numériques. Nous nous sommes montrés attentifs à leurs questionnements méthodologiques et épistémologiques, de même qu'aux dispositifs de recherche qu'ils ont pu inventer pour s'approcher au mieux des manières de s'informer. Nous avons effectué également une revue de littérature en sciences sociales à partir des titres suggérés d'abord par nos enquêtés, cette recherche bibliographique s'étant ensuite développée au fil de nos propres découvertes. Notre visée étant d'abord épistémologique et méthodologique, nous n'avons pas cherché à construire un « catalogue » complet des méthodes et techniques d'enquêtes. Il s'agissait plutôt de penser les stratégies méthodologiques dans leur rapport à des problématiques de recherche, en particulier celles concernant les manières de s'informer dans la société contemporaine. Ce rapport présente une synthèse analytique des résultats de cette « métaenquête ». Nous souhaitons que cette démarche exploratoire puisse permettre à des chercheurs, dans le cadre d'éventuels travaux dans le domaine, d'adapter leurs approches méthodologiques au nouvel environnement numérique et aux habitudes changeantes des individus et des publics en matière d'information.

Au chapitre II, nous rendons compte des problèmes et des enjeux les plus récurrents dans les propos des chercheurs que

---

2. Afin d'assurer la plus grande cohérence à notre travail d'enquête, nous avons choisi de nous concentrer sur le domaine des sciences sociales. Ainsi, nous avons exclu *a priori* à la fois le domaine des sciences de l'information et de la documentation, de même que les approches de type industriel mobilisées par les entreprises spécialisées dans la mesure d'auditoire (*user-centric*, *site-centric*, etc.) ou par les organisations médiatiques.

nous avons interrogés. Cette mise en évidence des problématiques et des enjeux permet d'identifier des repères susceptibles de nous orienter dans les discussions méthodologiques actuelles. Une fois décrite la complexification actuelle des pratiques informationnelles, nous nous intéressons aux inventions et suggestions méthodologiques des chercheurs, ce qui permet notamment de reconsidérer certains des axes d'analyse à privilégier pour faire avancer la recherche. Parmi les enjeux épistémologiques soulevés par nos enquêtés, les questionnements autour d'un possible « tournant computationnel » en sciences sociales nous semblent centraux parce qu'ils sont au cœur du développement des sciences sociales. Il s'avère en effet possible, aujourd'hui, de construire de très grands corpus de données (*big data*) du fait de l'existence d'internet et de ses dispositifs de traçage des trajectoires des usagers dans l'univers numérique.

Le chapitre III dresse un portrait des principaux types de méthodes de recherche en sciences sociales en lien avec le numérique : méthodes conventionnelles qualitatives et quantitatives ; ethnographies en ligne (dites « virtuelles ») ; méthodes computationnelles exploitant les très grands corpus de données (*big data*) ; méthodes numériques (*digital methods*) qui appellent à l'invention de nouvelles manières de faire de la sociologie en s'appuyant sur les données natives d'internet ; méthodes numériques quali-quantitatives qui articulent de manière renouvelée ces deux volets classiques des méthodes. Par-delà cette description des idéaux-types des méthodes, nous cherchons à identifier les principaux enjeux (organisationnels, économiques, politiques, éthiques) soulevés par cette variété de choix méthodologiques. D'une part, la question de la collaboration entre chercheurs en sciences sociales et informaticiens ; et d'autre part, les contraintes politiques et économiques dans l'accès aux traces des *big data* (aux fins de la construction des très grands corpus de données) nous semblent deux enjeux majeurs pour le développement contemporain de la recherche.

Le chapitre IV constitue le pivot de notre démarche d'enquête. Au terme d'un long processus de recherche et de consultation auprès de nombreux collègues, nous avons finalement choisi cinq études de cas correspondant aux idéaux-types retenus dans notre cartographie des méthodes. Chaque étude de cas concerne spécifiquement les manières de s'informer à l'heure du numérique.



Nous décrivons d’abord deux études faisant usage de méthodes conventionnelles en sciences sociales: d’abord, une étude quantitative des usages des médias numériques, ensuite une étude qualitative sur la lecture de la presse en ligne. La troisième étude de cas concerne une ethnographie en ligne des usages sociopolitiques de Facebook. La quatrième étude de cas retenue porte sur une étude sur le partage d’information par Twitter s’appuyant sur l’analyse d’un très grand corpus de Tweets et Retweets. La cinquième étude de cas présente une étude quali-quantitative<sup>3</sup> sur les usages de Facebook où les dimensions qualitative et quantitative sont enchevêtrées avec beaucoup de cohérence. Notre présentation de ces cinq études de cas est systématique dans la mesure où nous avons appliqué à chacune des études la même grille d’analyse: après une présentation synthétique de l’étude, nous cherchons à mettre en évidence les présupposés épistémologiques de la démarche; puis nous décrivons les assises théoriques sur lesquelles s’ancre la recherche; nous détaillons ensuite les techniques de collecte et d’analyse des données utilisées; subséquentement, nous présentons le type de résultats obtenus; enfin, nous discutons des limites du dispositif méthodologique mis en place dans chacune des études de cas.

Le chapitre V (Conclusions) présente une analyse comparative des diverses études de cas à l’aune d’un ensemble de critères scientifiques et épistémologiques. Nous terminons cette conclusion en présentant au CEM un ensemble de recommandations destinées aux chercheurs et chercheuses intéressés par le design de nouveaux dispositifs méthodologiques destinés à bien saisir les manières de s’informer des personnes et des publics, et s’avérant le mieux adaptés à l’environnement informationnel contemporain.

---

3. L’expression « méthodes mixtes » (*mixed methods*) est parfois utilisée aujourd’hui pour rendre compte de ces stratégies à double volet qualitatif et quantitatif. Toutefois, cette expression est trop générale car elle peut désigner indifféremment plusieurs assemblages de méthodes.



# 2

## Analyser les pratiques informationnelles dans un contexte d'instabilité :

### Défis épistémologiques et méthodologiques

La numérisation généralisée des multiples sphères de la vie sociale suscite nombre de discussions parmi les chercheurs en sciences sociales (Boullier, 2016). Sans même nous prononcer sur la pertinence des discours annonçant, du fait même de ce processus, une rupture dans le cours de l'histoire, force est néanmoins de constater, dans le champ de la recherche, un renouvellement des objets d'études, des altérations significatives dans les débats théoriques ou épistémologiques, mais aussi des déplacements dans le traitement des questions méthodologiques. Ce chapitre souhaite sensibiliser les lecteurs à ces interrogations, en présentant les problèmes et les enjeux les plus récurrents, sinon les plus centraux, dans les propos de chercheurs s'intéressant aux aspects méthodologiques de l'étude des activités en ligne et, plus spécifiquement, des pratiques informationnelles des usagers des

médias numériques<sup>4</sup>. Par cette démarche, nous n'entendons ni dresser une cartographie complète des questions agitant aujourd'hui la communauté scientifique ni apporter des réponses fermes et définitives à ces dernières. Il s'agit plutôt d'acquérir des repères pour être en mesure de nous orienter dans les discussions méthodologiques actuelles. Disons-le encore plus clairement : la visée de notre argumentation est exploratoire.

À cet effet, nous avons réalisé 24 entretiens, entre juin et novembre 2017, avec des chercheurs en sociologie, en communication et en gestion<sup>5</sup>. Les champs d'expertise de ces chercheurs francophones, situés au Québec ou en France, concernent la sociologie du numérique, le journalisme, la sociologie des médias, la sociologie des publics, la sociologie des réseaux sociaux, les études des sciences et des technologies (*Science and Technology Studies*), les humanités numériques (*Digital Humanities*), la sociologie de la jeunesse ou encore, la sociologie de la culture. Quoique travaillant sur une grande variété de phénomènes empiriques, à partir de perspectives théoriques le plus souvent distinctes, ces chercheurs ont tous été impliqués dans la conception de dispositifs méthodologiques visant à collecter et à traiter des données sur les pratiques informationnelles d'un ensemble déterminé d'acteurs. Leurs expériences dans ce domaine et, plus encore, leur réflexivité par rapport à ces expériences, se sont révélées précieuses pour appréhender les coordonnées des débats animant aujourd'hui la communauté scientifique au sujet des méthodes de recherche sur les pratiques informationnelles. Pour obtenir des regards en contrepoint des discours universitaires, nous avons également réalisé un entretien avec un membre du bureau de direction de la Fédération Internet

---

4. Les pratiques informationnelles dont il est question s'inscrivent dans le contexte de démocraties libérales. Cette précision mérite d'être formulée, dans la mesure où la prise en compte des conditions de production, de circulation et de réception de l'information dans des régimes autoritaires déplacerait substantiellement les coordonnées de notre analyse.

5. La liste des personnes interrogées se trouve dans l'annexe de ce rapport.

Nouvelle Génération<sup>6</sup> (FING), d'une part, et un entretien avec un entrepreneur ayant une expertise dans le domaine du « marketing des données<sup>7</sup> », de l'autre.

Il s'agira, dans un premier temps, de s'interroger sur les évolutions actuelles des pratiques informationnelles dans l'idée de cerner ce qui, dans ce contexte, requiert de penser à nouveaux frais les méthodes de recherche en sciences sociales sur les manières de s'informer. Or, de ce point de vue, plusieurs éléments affectent le travail des chercheurs. La diversification des sources d'information, la fragmentation de la consommation médiatique, la multiplication des supports matériels, la généralisation du *multitasking* (pluriactivité : tendance, parmi les récepteurs, à mener plusieurs tâches simultanément), ainsi que l'instabilité des pratiques informationnelles, sont en effet autant de facteurs contribuant à l'opacité de ces phénomènes, constituant par là même de véritables défis pour les chercheurs en sciences sociales. Si ce contexte extraordinaire exige minimalement une réévaluation rigoureuse des modalités méthodologiques de l'analyse des pratiques informationnelles, une partie de la communauté scientifique fait d'ores et déjà preuve d'ingéniosité et d'inventivité, d'une part, en réajustant des techniques d'enquête classiques, et d'autre part, en expérimentant de nouvelles approches et de nouveaux dispositifs d'enquête susceptibles d'épouser les contours des pratiques informationnelles actuelles.

Partant de là, nous passerons rapidement en revue les aspects les plus fréquemment mentionnés dans nos entretiens au sujet

---

6. La Fédération Internet Nouvelle Génération (FING) est un groupe de réflexion (*think tank*) spécialisé dans les transformations numériques. Cette association, créée en 2000, rassemble des entreprises, des laboratoires de recherche, des universités, des professionnels, des personnes physiques, des administrations publiques, mais aussi des collectivités territoriales. FING. (2017). Présentation. Repéré à <http://fing.org/?Presentation&lang=fr>.

7. Ce terme désigne « toutes les décisions et actions marketing basées sur les données relatives aux clients ou prospects ou sur d'autres types de données liées à l'environnement [...]. Le marketing de la donnée occupe une place de plus en plus importante dans le domaine du marketing avec l'utilisation croissante des données qui est faite dans le domaine du marketing digital et avec la montée en puissance du Big Data. » Définitions marketing. (2017). Marketing des données. Repéré à <https://www.definitions-marketing.com/definition/marketing-de-la-donnee/>

de ces « bricolages méthodologiques ». Nous examinerons d'abord ce qu'implique cette évolution de notre environnement médiatique au regard des axes analytiques des recherches sur les pratiques informationnelles. Nous rendrons également compte des préoccupations des personnes que nous avons interrogées au sujet des catégories de données légitimement exploitables pour conduire des recherches sur les manières de s'informer des individus. Nous prêterons enfin attention à la question de la pertinence d'une hybridation – d'aucuns parleraient de « *mixed methods* » – des méthodes quantitatives (autant les méthodes computationnelles que les méthodes statistiques) et qualitatives pour nous adapter aux spécificités du contexte actuel.

Nous nous arrêterons enfin sur les considérations épistémologiques les plus prégnantes dans nos entretiens. Nous reprendrons alors à notre compte des arguments s'inscrivant en faux contre la réduction – quelque peu naïve – des choix méthodologiques à de simples « outils ». Les discours des 24 chercheurs témoignent, à n'en pas douter, d'interrogations techniques sur les méthodes à mettre en œuvre pour appréhender les pratiques informationnelles, mais aussi d'une conscience aiguë des enjeux épistémologiques accompagnant la production de connaissances à partir de ces mêmes méthodes. Voilà pourquoi nous aborderons les questions de la nature des savoirs émanant de ces dernières et de l'idée d'une rupture possible dans le développement historique des sciences sociales.

## **1. Une complexification des pratiques informationnelles**

Pour donner à nos entretiens une intelligibilité pleine et entière, permettons-nous au préalable la formulation de quelques observations concernant les particularités de l'évolution des pratiques informationnelles depuis un peu plus d'une décennie. Ces dernières possèdent en effet plusieurs propriétés dont la connaissance demeure indispensable pour apprécier les enjeux et les problèmes méthodologiques de la recherche, en sciences sociales, portant sur ces phénomènes. Si nos entretiens ne nous autorisent pas à établir un tableau d'ensemble sur les conditions

actuelles de la production, de la circulation et de la réception de l'information, ils comportent suffisamment d'éléments pour nous permettre d'entrevoir les contours empiriques des pratiques informationnelles aujourd'hui. Or, ce qui ressort des discours de nos acteurs, c'est d'abord le constat d'une complexification – parfois vertigineuse et désarmante – des manières de s'informer.

*La diversification des sources d'information* apparaît d'emblée comme un facteur déterminant de complexification. Les usagers bénéficient maintenant de la possibilité de créer, de partager et de consulter des contenus par l'entremise des médias socionumériques<sup>8</sup>, ceux-ci pouvant prendre la forme de plateformes de partage de contenus audiovisuels (YouTube, Dailymotion), de réseaux sociaux (Facebook, Google+), de plateformes de micro-bloggings (Twitter) et de bloggings (Tumblr) ou encore de sites de partage de photographies (Instagram, Flickr). L'émergence de nouveaux acteurs dans le champ médiatique vient également accentuer cette tendance à la diversification des sources d'information. Aux médias traditionnels, qui travaillent selon des modalités variables à la coordination de leurs activités en ligne et hors ligne, s'ajoutent les plateformes des médias n'ayant plus qu'une présence en ligne (les *pure players*), mais aussi les médias (ayant des équipes éditoriales bien dotées et bien organisées) dont l'activité principale consiste à identifier et à diffuser des contenus qui sont « dans l'air du temps » et susceptibles de faire le « buzz » sur Internet<sup>9</sup>. Il faut également compter avec l'introduction très récente (2016-2018), dans le champ médiatique, d'un nouveau type d'acteurs ayant la particularité de n'entretenir aucune plateforme d'information, d'une part, et de diffuser leurs propres contenus exclusivement par l'intermédiaire de médias socionumériques déjà existants, d'autre part<sup>10</sup>. Les productions de ces nouveaux acteurs – par exemple, en France, ces jeunes entreprises ont pour noms : Brut., Explicite, TF1 One, Monkey, Loopsider – empruntent la forme de vidéos d'information

---

8. Dans notre analyse, nous utiliserons les expressions « médias socionumériques » et « plateformes numériques » comme des synonymes.

9. BuzzFeed étant probablement l'organisation la plus populaire dans cette catégorie d'acteurs médiatiques.

10. Par exemple, le média « Brut. » ne diffuse ses contenus – des vidéos – que par l'intermédiaire de différents réseaux sociaux : Facebook, Youtube, Twitter, Dailymotion, Snapchat et Instagram.

(surtout des formats courts, mais pas seulement : il y a aussi un public qui demande des vidéos longues) destinées surtout aux jeunes (15-25 ans) fréquentant assidûment les réseaux sociaux numériques. Au Québec, ce sont des médias traditionnels qui font des expériences similaires. Elles ont pour nom Rad (à Radio-Canada) ou Tabloid (au Journal de Montréal), par exemple. Selon ces nouveaux acteurs de l'information vidéo, qui se développent selon des lignes éditoriales pourtant assez contrastées, le consensus serait que l'avenir de l'information passera par les réseaux sociaux et leurs abonnés d'abord identifiés à la catégorie marketing des « milléniaux » (*Millennials*: 15-25 ans). Ces jeunes identifient les médias sociaux comme leur source principale d'information (Reuters, 2017). Ils manifestent le besoin de comprendre mieux le monde dans lequel ils vivent. Ils souhaitent répondre à ce besoin de manière synthétique, divertissante et même parfois humoristique, sous un angle nouveau et inédit au regard des médias traditionnels, ce qui ne les empêche pas de manifester une envie de s'engager dans des activités de type social ou communautaire. En conséquence, il s'agit pour les nouveaux acteurs producteurs de l'information vidéo de se montrer suffisamment attrayants et innovants pour pouvoir capter l'attention de ces jeunes clientèles. Et, en même temps, de réussir à inventer un modèle économique qui s'avérera rentable à moyen et à long terme (Thomann, 2017).

Cette diversification de l'offre informationnelle s'accompagne d'une *fragmentation de la réception*. Nos entretiens témoignent en effet de l'épuisement progressif d'un mode de consommation médiatique fondée sur l'exposition, plus ou moins durable, des usagers à quelques sources d'information. Bien que cette affirmation fasse l'objet de nuances importantes – les modes de consommation médiatique pouvant varier en fonction de la position des individus dans l'espace social, des catégories d'âge ou de paramètres culturels – force est de constater des familiarités entre nos entretiens sur ce point, sinon des convergences. Ceux-ci corroborent l'idée d'une consommation médiatique fragmentée, passant par la fréquentation d'un nombre grandissant de sources



d'information<sup>11</sup>. Cela signifie non pas que nous assistions à « une révolution » du pôle de la réception dans le processus de circulation de l'information, mais que ce phénomène, déjà constaté de longue date dans les recherches sur la réception dans les médias traditionnels comme la télévision (Proulx et Maillet, 1998), se renforce avec l'entrelacement des médias traditionnels et numériques. À la diversification de l'offre informationnelle semble alors correspondre une fragmentation accentuée de la consommation médiatique<sup>12</sup>.

*La multiplication des supports matériels* contribue à son tour à ce processus, dans la mesure où les usagers s'informent désormais au moyen de différentes technologies de communication numériques (*smartphones*<sup>13</sup>, tablettes, ordinateurs de bureau, portables, objets connectés, etc.), tout en continuant à consulter la presse écrite, à écouter la radio ou à regarder la télévision. La pluralité des supports et des médias se révèle être, de ce fait, une composante à considérer dans l'analyse des manières de s'informer aujourd'hui.

- 
11. Bien que cette idée ne soit pas explicitement évoquée dans nos entretiens, l'hypothèse d'une accentuation de la fragmentation de la réception trouve également des confirmations dans les recherches sur la sérendipité dans les recherches d'information en ligne (Andel et Bourcier, 2009). La sérendipité désigne, en l'occurrence, la capacité à faire des découvertes aléatoires en « régime d'exploration curieuse » (Auray, 2016). Bref, la disponibilité et la capacité de « trouver ce que l'on ne cherche pas ».
  12. Par ailleurs, cette fragmentation de la consommation médiatique s'accompagne parfois de l'utilisation de services ou de logiciels d'agrégation de contenus (Feedly), offrant aux usagers la possibilité de réunir dans une seule et même interface des informations provenant de différentes sources. Les réseaux sociaux assument une fonction similaire, en autorisant les usagers à rassembler un nombre pléthorique de sources d'information sur leur compte.
  13. Au Québec, on parle de « téléphone intelligent ».

*La généralisation de la pluriactivité (multitasking)* participe aussi à l'opacité des pratiques informationnelles contemporaines<sup>14</sup>. Nos entretiens mettent effectivement en lumière une tendance, dans le comportement des récepteurs, à mener plusieurs tâches simultanément (répondre à des courriels avec un téléphone intelligent en attendant un bus, visualiser des contenus divertissants sur un ordinateur portable pendant un cours, consulter son compte Facebook pendant une réunion, etc.). Cet entrelacement des pratiques constitue un défi redoutable pour les chercheurs, car il fait obstacle à la possibilité d'inférer une activité cognitive effective en se basant seulement sur les traces des usagers en ligne, d'une part, et parce qu'il compromet la fiabilité de la mémoire des usagers au regard de leurs pratiques informationnelles, d'autre part.

Pour terminer, nos entretiens ont témoigné à plusieurs reprises de vives préoccupations au sujet de *l'instabilité dynamique des pratiques informationnelles*. Ces dernières semblent en effet évoluer à un rythme très rapide, sous des formes multiples, dessinant ainsi des tendances peu prévisibles par les experts du domaine. Plusieurs chercheurs nous ont d'ailleurs expliqué leur sentiment d'incertitude face à des phénomènes susceptibles de connaître des changements majeurs entre le début et la fin de la réalisation d'une même étude. L'ambition de décrire, d'expliquer, de mesurer ou de comprendre les manières de s'informer apparaît ainsi pour le moins compromise par leur volatilité. La question de la possibilité de rendre compte, au-delà d'une photographie ponctuelle, de pratiques informationnelles aussi éphémères se

---

14. Cette catégorie de *multitasking* (pluriactivité) – « le fait de devoir faire plusieurs choses en même temps, d'être sollicité sans répit » (Datchary, 2011) – ou de « dispersion », a été développée dans un contexte de sociologie du travail qui l'associe à une possible surcharge attentionnelle liée à une injonction à la performance et à une souffrance au travail (Datchary, 2011). Nous reprenons cette catégorie ici pour décrire la pluriactivité des personnes agissant dans un écosystème informationnel caractérisé par une abondance de l'offre des messages, provoquant une surcharge attentionnelle. En même temps, la capacité à la pluriactivité devient une compétence pour bien circuler dans le monde numérique.

pose ainsi dans toute son acuité<sup>15</sup>. Cette dimension temporelle concourt là encore au processus de complexification des manières de s'informer.

## 2. Le « bricolage » méthodologique

Les observations précédentes, au regard de la complexification des pratiques informationnelles, engendrent naturellement un doute sur ce que nous pouvons espérer savoir de ces dernières, en tant que chercheurs en sciences sociales. Quelles stratégies et tactiques méthodologiques, dans le contexte d'une diversification des sources d'information, d'une fragmentation de la consommation médiatique, d'une multiplication des supports matériels, d'une généralisation du *multitasking* et de l'instabilité des pratiques informationnelles, pouvons-nous en effet imaginer pour appréhender des phénomènes dont les contours empiriques sont aussi irréguliers que variables? Ce problème, omniprésent dans nos entretiens, n'a suscité que peu de réponses définitives. Alors qu'un petit nombre de chercheurs ont affirmé leur confiance dans des dispositifs méthodologiques mixtes qu'ils testent depuis plus d'une dizaine d'années, la plupart des acteurs que nous avons interrogés privilégiaient volontiers le champ lexical du « bricolage » pour exposer les méthodes de recherche dans ce domaine<sup>16</sup>.

Pour éviter d'inutiles équivoques, précisons d'emblée la signification de ce champ lexical dans nos entretiens. Loin de porter un sens péjoratif – le terme « bricolage » pouvant renvoyer à une démarche d'essais et erreurs, qui serait à l'opposé d'une procédure rigoureuse et systématique –, ce vocabulaire venait en réalité souligner à la fois l'indétermination, l'inventivité permanente et

---

15. Il faudrait aussi mentionner, pour prendre pleinement la mesure de cette instabilité, les inflexions des pratiques informationnelles en fonction de l'âge des individus. Dans le contexte de leurs travaux, plusieurs chercheurs ont effectivement réalisé que les manières de s'informer de leurs enquêtés, en l'occurrence des adolescents, avaient évolué significativement en l'espace de quelques mois.

16. Des expressions telles que « bricolage », « essai », « artisanal » ou « à la main » étaient fréquemment employées par les personnes interrogées pour caractériser le travail de construction de dispositifs méthodologiques qu'elles réalisent dans leurs recherches.

la richesse des stratégies méthodologiques dans les travaux actuels sur les activités en ligne et, *a fortiori*, sur les manières de s'informer des individus. La métaphore de la « boîte noire », quelquefois employée dans nos entretiens, permettait en l'occurrence de se représenter adéquatement ce qui caractérise la phase dans laquelle nous nous trouvons aujourd'hui : une phase d'ouverture aux expérimentations méthodologiques<sup>17</sup>. Pour le dire autrement, le bricolage désigne, dans notre argumentation, non pas un travail amateur et bâclé, mais une activité témoignant au contraire d'une « ingéniosité adroite, valorisée » (Rey, 2012) dans un contexte pour le moins changeant. Force est par ailleurs de constater qu'il y a des disparités dans la communauté scientifique, les acteurs n'ayant pas tous le même degré d'expérience dans les recherches sur les pratiques informationnelles. La prudence exige par conséquent de ne pas étendre ce constat du bricolage à l'ensemble des recherches menées dans ce domaine.

## 2.1 Repenser les axes d'analyse

En brossant le portrait d'un paysage médiatique composite et versatile, notre panel d'experts et de chercheurs a été naturellement amené à mettre en question les axes d'analyse à privilégier dans l'étude des pratiques informationnelles actuelles. Cette interrogation s'est vue en outre renforcée par l'hypothèse, communément partagée par ces derniers, d'une reconfiguration par les réseaux sociaux – et notamment par la plateforme de Mark Zuckerberg – des modalités par lesquelles les individus s'informent aujourd'hui. Bien qu'ils ne défendent pas une conception déterministe de la communication médiatique, selon laquelle l'évolution technologique engendrerait mécaniquement et unilatéralement des changements sur les plans de la production, de la diffusion et de la réception de l'information, nos entretiens font néanmoins apparaître l'émergence des réseaux socionumériques comme une condition nécessaire, mais insuf-

---

17. Cette métaphore (d'abord issue de la cybernétique) s'inspire de la théorie de l'acteur-réseau et, plus spécifiquement, des travaux de Bruno Latour (2001).

fisante, des mutations que connaît actuellement la communication médiatique<sup>18</sup>.

Plusieurs acteurs de notre panel ont insisté sur la centralité, parfois même sur la primauté, de Facebook dans notre environnement médiatique<sup>19</sup>. Dans la mesure où les récepteurs agrègent et consultent de multiples contenus par cette plateforme, celle-ci se situerait, aujourd'hui, au cœur des processus de circulation de l'information. Cela induirait alors une réarticulation des rapports entre les organisations médiatiques et les récepteurs, ces rapports étant encadrés dans des relations sociales de « partage<sup>20</sup> ». Les pratiques informationnelles se déploieraient ainsi non seulement dans l'exposition directe des récepteurs aux sources d'informations des organisations médiatiques, mais aussi dans des interactions médiatisées entre les usagers des réseaux sociaux.

Ce contexte, marqué par une complexification et une reconfiguration des pratiques informationnelles, exige de repenser les axes d'analyse des recherches en sciences sociales sur les manières de s'informer. Or, reconsidérer ces dernières ne consiste pas à faire « tabula rasa ». Il s'agit plutôt d'évaluer la pertinence des axes d'analyse communément adoptés dans la littérature scientifique sur l'étude des médias, dans l'idée d'envisager la possibilité de les ajuster aux altérations de notre paysage médiatique, de les conserver tels quels ou d'en

---

18. Des facteurs politiques, économiques, culturels, démographiques et sociaux affectent également ce processus de mutation de la communication médiatique. La théorie interactionnelle des médias de communication, élaborée par John B. Thompson, demeure une option pertinente pour conceptualiser l'influence de l'évolution technologique sur les diverses sphères de la vie sociale sans verser dans le déterminisme (1995). Il est nécessaire de bien distinguer le poids des « déterminations » (catégorie essentielle des sciences sociales) de la perspective du « déterminisme » qui suppose une exclusivité dans l'explication causale par la catégorie mise de l'avant.

19. Cette analyse est corroborée par les récentes études sur le taux de pénétration des médias socionumériques ([https://www.perspectives-communication.com/newsroom/communication-creation-medias-blog\\_files/chiffres-internet-reseaux-sociaux.php](https://www.perspectives-communication.com/newsroom/communication-creation-medias-blog_files/chiffres-internet-reseaux-sociaux.php))

20. Le terme « partage » mérite d'être employé avec prudence, dans la mesure où il n'est pas dépourvu de dimensions idéologiques. L'expression « économie du partage », associée aux plateformes numériques sociales (Facebook, Twitter, etc.), de services (Uber, AirBnB, etc.) ou encore de sociofinancement (Kickstarter) devrait suffire à susciter un doute légitime à ce sujet (Proulx, 2018).

développer de nouveaux. D'après les personnes que nous avons interrogées, plusieurs dimensions nécessitent d'être intégrées pour pouvoir épouser les contours des pratiques informationnelles actuelles :

- i. La connaissance des *sources de l'information*. D'où proviennent les informations ? Comment qualifier les acteurs collectifs ou individuels à l'origine des informations ?
- ii. La connaissance des *modalités d'accès aux informations*. Par quels dispositifs sociotechniques<sup>21</sup> les individus accèdent-ils à l'information ?
- iii. La connaissance des *modalités de « partage » de l'information* entre pairs. Comment l'information circule-t-elle entre les individus ?
- iv. La connaissance des *modalités d'attribution d'une valeur sociale* aux informations par les individus. Selon quelles logiques sociales les individus catégorisent-ils les informations et leur accordent-ils de la crédibilité<sup>22</sup> ?
- v. La connaissance des *caractéristiques socioéconomiques des individus eux-mêmes*. Quelle place les individus occupent-ils dans l'espace social ?

## 2.2. Les phénomènes d'influence dans les processus de circulation de l'information

Nous aimerions revenir sur le troisième axe d'analyse, dans la mesure où celui-ci permet – peut-être mieux que les autres – d'entrevoir ce qui se joue derrière ces déplacements analytiques. L'étude des pratiques informationnelles requiert au minimum d'examiner trois aspects des modalités de « partage » de l'infor-

---

21. « Un dispositif sociotechnique pour l'information et la communication est une mise en œuvre spontanée ou programmée, dans laquelle des acteurs "jouent" avec des ordinateurs en réseau pour chercher de l'information, transmettre de l'information, échanger ou communiquer, etc. » (source : [https://fr.wikipedia.org/wiki/Dispositif\\_sociotechnique\\_pour\\_l%27information\\_et\\_la\\_communication](https://fr.wikipedia.org/wiki/Dispositif_sociotechnique_pour_l%27information_et_la_communication)).

22. L'analyse des modalités d'accès aux contenus informationnels ne saurait en effet être réduite au recensement des dispositifs sociotechniques par lesquels les individus s'informent. Cette analyse suppose par exemple de s'intéresser aux logiques de prescription et aux modalités de construction de la confiance, processus par lesquels les récepteurs accordent une crédibilité aux contenus.

mation : les propriétés des relations sociales entre les individus ; les caractéristiques des informations partagées par les individus ; et les phénomènes de rétroaction dans leurs échanges (commentaires, rediffusions, etc.)<sup>23</sup>. L'intégration de ces différents aspects ne date évidemment pas du 21<sup>e</sup> siècle dans les recherches sur la communication médiatique. Il suffit de penser, par exemple, aux contributions de l'École de Columbia<sup>24</sup> – notamment l'hypothèse confirmée du *two-step flow* ou le courant des *Uses and Gratifications* – pour s'en convaincre (Blumler et Katz, 1974). Cela étant dit, la numérisation de l'environnement médiatique accentue l'importance des processus de circulation de l'information entre les individus et, possiblement, celle de leur capacité d'influence interpersonnelle. Il s'agit certainement d'une piste de recherche à explorer dans des travaux futurs.

Plusieurs répondants de notre panel estiment ainsi nécessaire de mettre l'accent sur les réseaux personnels d'influence dans les pratiques informationnelles. Quoiqu'indéterminée dans nos entretiens, la notion d'« influence » n'est évidemment pas sans lien avec le niveau d'activité des acteurs, leur visibilité ou encore leur aptitude à provoquer des rétroactions. Les recherches en sciences sociales sur l'influence personnelle sortiraient alors renforcées de ce contexte, dans lequel priment quelques individus jouissant d'une position centrale dans les relations sociales d'échange et de partage. La revue de littérature sur les travaux empiriques à la source de la définition et du développement contemporain du « marketing viral » revêt ici un intérêt certain (Mellet, 2009). Le poids de l'influence personnelle varie selon les situations envisagées, le leadership ne s'exerçant pas nécessairement de manière unidimensionnelle et unidirectionnelle. Ce constat n'est pas sans rappeler les travaux pionniers d'Elihu Katz et Paul Lazarsfeld (1955/2006) sur les « leaders d'opinion » influençant les personnes qu'ils et elles côtoient en reprenant et réinterprétant certaines informations choisies, circulant dans les

---

23. Cela n'implique toutefois pas de restreindre le champ de l'analyse aux seules interactions par les plateformes en ligne. Les processus de circulation de l'information se déploient en effet autant en ligne qu'hors ligne (discussions sur les lieux de travail, à table, dans les transports en commun, etc.) et sont d'ailleurs articulés entre eux (Boullier, 2004).

24. Pour une description détaillée des contributions de l'École de Columbia, voir Breton et Proulx, 2012.

médias. Par ailleurs, d'un point de vue méthodologique élargi, la norme de la « représentativité » dans la construction des échantillons pour fins d'enquêtes par sondage en sciences sociales – qui postule qu'un échantillon bien construit a la capacité de « parler pour la population » – se voit bousculée par l'importance prise par le Web et les médias sociaux dans le paysage médiatique contemporain. Du point de vue de l'évolution des connaissances sur les publics médiatiques, l'un des basculements épistémologiques majeurs s'ancre dans le fait que l'on serait passé « d'une logique de la représentativité à une logique de l'influence » (Beaudouin et Denis, 2014, p. 25). L'une des chercheurs interrogés a ainsi fait référence aux travaux du jeune sociologue Baptiste Kotras (2013) enquêtant sur les firmes engagées dans la promotion de l'e-réputation et qui remettent en question les manières habituelles d'échantillonner des populations dans le nouveau contexte numérique. Kotras trace la distinction entre l'échantillon représentatif (pour un traitement statistique classique du sondage d'opinion, par exemple) et l'échantillon construit au moyen de critères alternatifs (influence, visibilité). Dans ce deuxième cas de figure sont choisies les « opinions qui comptent », c'est-à-dire celles qui ont « du poids » en termes d'influence et de visibilité. Cette façon d'échantillonner apparaît plus adaptée à l'environnement propre aux réseaux socionumériques, puisqu'elle prend directement en compte la capacité de ces réseaux à générer une plus grande visibilité des agents « influents ».

### **2.3 Quelles données employer pour étudier les pratiques informationnelles actuelles ?**

Les « bricolages méthodologiques » ne relèvent pas seulement de l'identification des axes analytiques à privilégier dans les recherches, en sciences sociales, sur les pratiques informationnelles. Nos entretiens comportent en effet de nombreuses réflexions au sujet des catégories de données ayant une pertinence dans l'étude des manières actuelles de s'informer. Or, à partir de la synthèse des discours des personnes que nous avons interrogées, quatre catégories de données peuvent être déclinées : les données collectées par les techniques d'enquête conventionnelles ; les mesures d'auditoire des entreprises spécialisées dans ce domaine ; les données fournies par les organisations médiatiques ; et les données produites par les plateformes en ligne.



Sans entrer pour l'instant dans les détails de ces stratégies méthodologiques, les enquêtés évoquaient d'abord la possibilité de collecter des données à partir d'une combinaison variable de techniques d'enquête bénéficiant, en règle générale, de la reconnaissance de la communauté scientifique<sup>25</sup>. Leur utilisation peut alors s'inscrire dans le cadre de démarches quantitatives, qualitatives ou mixtes. Notons que ces techniques d'enquête se voyaient souvent attribuer les qualificatifs « traditionnelles », « conventionnelles » ou « classiques » dans les discours de notre panel de chercheurs et d'experts. Loin d'être anecdotique, le recours à ces adjectifs témoigne d'un questionnement sur la réalité d'une discontinuité, sinon d'une rupture, dans l'évolution des méthodes de recherche en sciences sociales. Nous reviendrons là-dessus dans la partie subséquente.

Il demeure aussi envisageable de recueillir des données auprès des entreprises spécialisées dans les mesures d'auditoire (Médiamétrie, Nielson, ComScore, etc.) ou offrant des services de ce type (Google Analytics). Quoique ces organisations privées soient critiquées pour l'absence de contextualisation ou le manque de précision de leurs statistiques, elles fournissent des données qui n'en sont pas moins utiles pour appréhender, même partiellement, les manières de s'informer des individus, autant d'ailleurs avec les médias traditionnels qu'avec les médias numériques. Les organisations médiatiques, parce qu'elles conservent et archivent une pléthore de données sur leurs auditoires (sur les caractéristiques démographiques ou socioéconomiques de leurs abonnés, sur les traces des utilisateurs de leurs plateformes en ligne, etc.), restent également de précieuses ressources pour cerner les contours des pratiques informationnelles aujourd'hui. Notre panel d'experts et de chercheurs a néanmoins exprimé des réserves par rapport à l'utilisation de ces données, les organisations médiatiques pouvant, en fonction de leurs intérêts économiques, interdire l'accès à ces dernières. De plus, à l'instar des statistiques du marché des mesures d'auditoire – hormis le cas de données produites dans le cas d'enquêtes spécifiques et financées en conséquence – les données des organisations médiatiques

---

25. Celles-ci sont fréquemment exposées dans les ouvrages de référence d'une discipline ou d'un domaine de recherche, dans les manuels, dans les cours des programmes de 1<sup>er</sup> cycle ou de cycles supérieurs, etc.

peuvent s'avérer trop décontextualisées ou générales pour des fins de recherche<sup>26</sup>.

Pour terminer, les médias socionumériques donnent accès à une multitude de données sur les activités en ligne et les profils de leurs utilisateurs. Ces données peuvent être collectées en exploitant l'interface de programmation (*Application Programming Interface – API*) des plateformes numériques ou en procédant au *scraping* de données. L'obtention de ces données s'effectue, par conséquent, soit en suivant les modalités telles qu'elles ont été définies par les concepteurs de ces plateformes sociales, soit en utilisant des techniques d'extraction à partir de logiciels dits « clients » (navigateurs web, etc.), installés directement sur les périphériques des usagers<sup>27</sup>. Un accord peut également être négocié entre les chercheurs et les responsables de ces plateformes sociales pour l'obtention d'un jeu déterminé de données.

Bien que nous entendions approfondir, dans le prochain chapitre, les enjeux méthodologiques de la collecte et du traitement des données des plateformes numériques, évoquons d'emblée plusieurs points revenant avec insistance dans nos entretiens. Nos enquêtés ont souligné la richesse de ces données, non sans attirer toutefois notre attention sur les limites que ces dernières comportent. Dans notre panel, personne n'a contesté les promesses que ces données offrent aux recherches, en sciences sociales, sur les pratiques informationnelles. Grâce à leur niveau de granularité, elles permettent d'analyser avec une grande précision les manières de s'informer d'un individu, et ce, dans des situations variables. En outre, leur volume autorise les chercheurs à envisager la production d'une connaissance très détaillée des activités en ligne pour un vaste ensemble d'utilisateurs. Ces données, qui combinent simultanément la finesse et la taille,

---

26. Au-delà des données qu'ils peuvent fournir sur les pratiques informationnelles, les médias et les organisations spécialisées dans les mesures d'auditoire possèdent une précieuse expertise dans la conception de dispositifs méthodologiques.

27. Le *scraping* de données ne requiert donc pas l'autorisation des propriétaires des plateformes sociales. Ceux-ci tolèrent plus ou moins ce procédé d'extraction, en fonction du niveau de sensibilité de leurs données et de leurs intérêts économiques. Cette technique révèle par là même l'existence d'un rapport de pouvoir, dont l'enjeu n'est autre que le droit d'accéder aux données des utilisateurs (Marres et Weltevrede, 2013).

s'avèrent par là même fécondes d'un point de vue méthodologique. Il n'en reste pas moins qu'elles supposent un travail de « nettoyage » non négligeable pour être exploitées. Plus encore, elles ne sont pas signifiantes en elles-mêmes. Pour apporter une quelconque intelligibilité aux pratiques informationnelles, les données des plateformes en ligne requièrent non seulement des opérations techniques (pour les corriger, les simplifier et les homogénéiser), mais aussi un effort de contextualisation. Dans cette perspective, les résultats de nos entretiens constituent une mise en garde contre l'idée selon laquelle nous disposerions, à présent, de données brutes pouvant être immédiatement mobilisées dans des études en sciences sociales.

## 2.4 Quelques définitions

L'argumentation précédente révèle une différence de nature entre les données des plateformes en ligne et les données produites par les chercheurs en sciences sociales. Les opérations de formalisation et de contextualisation participent ainsi d'un processus de transformation, visant à assurer l'adéquation des données des plateformes en ligne aux normes et objectifs de la recherche en sciences sociales. Or, d'après les chercheurs que nous avons interrogés, cette « traduction » de l'ingénierie informatique aux sciences sociales peut être effectuée en travaillant à l'hybridation des méthodes qualitatives et quantitatives. La mixité de ces approches ne saurait garantir le succès de l'appropriation des données informatiques par la communauté scientifique. Elle apparaît toutefois, dans nos entretiens, comme une piste à explorer pour qui souhaite entreprendre l'analyse des pratiques informationnelles à partir des données des plateformes numériques<sup>28</sup>.

En sciences sociales, les méthodes qualitatives s'identifient à l'ensemble des procédés par lesquels les chercheurs recueillent et traitent des données permettant de saisir, tant sur le plan individuel que sur le plan collectif, les expériences vécues des

---

28. Dans cette sous-partie, nous sommes dans la nécessité de prendre un peu plus de distance par rapport au contenu de nos entretiens. Cette démarche est requise pour définir un certain nombre de notions, sans lesquelles les réflexions de notre panel de chercheurs et d'experts resteraient obscures.

acteurs ou de décrire leurs pratiques sociales en les réinscrivant dans leur contexte. Une approche qualitative peut conjuguer plusieurs techniques d'enquête, comme les entretiens individuels, les groupes de discussion (*focus groups*) ou encore les observations ethnographiques. Les méthodes quantitatives susceptibles d'être croisées aux méthodes qualitatives appartiennent quant à elles à deux catégories: les méthodes statistiques<sup>29</sup> et les méthodes computationnelles. Si le recours aux statistiques accompagne la naissance des sciences sociales à la fin du 19<sup>e</sup> siècle – les données statistiques sont des instruments d'objectivation qui permettent de dissocier le niveau des représentations individuelles d'avec le niveau social ou collectif (Martin, 1999)<sup>30</sup>; la statistique constitue ainsi la « première pierre de la sociologie » (Héran, 1984) –, la situation apparaît quelque peu différente pour les méthodes computationnelles.

Se développant en sciences sociales à partir de la fin des années 1990, les utilisations computationnelles de l'ordinateur existent en physique, en chimie et en sciences de l'ingénieur dès les années 1980 (Varenne, 2010, p. 17). Or, « un usage computationnel [...] fait de l'ordinateur une machine à modéliser au moyen de computations » et « une computation est une opération pas à pas sur des symboles. Elle peut elle-même être interprétée comme un calcul de modèle ou comme la simulation directe du comportement d'un système cible » (Varenne, 2010, p. 18). D'après ces définitions successives, les méthodes computation-

---

29. La statistique peut être définie, au sens le plus général, comme l'ensemble des méthodes et des techniques permettant de produire des mesures à partir de données provenant d'un échantillon (Dodge, 2007). Voir le texte instructif et convaincant de Héran (1984) pour saisir l'importance de la statistique dans la fondation même de la sociologie.

30. « Offrant une représentation du monde social extérieur aux subjectivités et perceptions individuelles [les statistiques] permettent de répondre à l'impératif méthodologique de Durkheim : pour étudier les faits sociaux, il faut les traiter comme des choses et, pour cela, 'il faut considérer les phénomènes sociaux en eux-mêmes, détachés des sujets conscients qui les représentent; il faut les étudier du dehors comme des choses extérieures' (Durkheim) » (Martin, 1999, p. 501).

nelles ont donc pour dessein soit le calcul d'un modèle, soit la simulation<sup>31</sup>.

Au sens général du terme, le « modèle » désigne à son tour « un objet, de nature symbolique ou matérielle, qui, dans le cadre d'un questionnement, a pour fonction minimale de faciliter un accès cognitif à un objet cible via une médiation d'un type donné » (Varenne, 2010, p. 20). Partant de là, cet « objet symbolique ou matériel » peut appartenir à trois grandes catégories : les modèles pour l'analyse de données, les modèles prédictifs et les modèles explicatifs. Les premiers servent à « faciliter une expérience ou la présentation de données d'expérience », les deuxièmes à « organiser de manière pratique et économe le matériau empirique (selon le principe du rasoir d'Ockham<sup>32</sup>) tout en permettant de faire des prédictions utiles » et les troisièmes à « tester des hypothèses de mécanismes opérant au niveau de l'individu » (Varenne, 2010, p. 24-25). Sur la base de ces distinctions, l'hybridation des méthodes dont parlaient nos enquêtés ne concernait pas l'ensemble des méthodes computationnelles, mais, plus particulièrement, celles employées pour l'édification de modèles d'analyse de données. Cette focalisation se comprend aisément au regard de l'importance acquise par les *big data* dans les réflexions méthodologiques et épistémologiques des chercheurs en sciences sociales aujourd'hui (Boullier, 2015a ; 2015b ; Boyd et Crawford, 2012 ; Kitchin, 2014).

Depuis quelques années, la communauté scientifique a la possibilité d'exploiter des jeux de données caractérisées à la fois par leur volume, leur vélocité, leur variété, leur exhaustivité, leur finesse, leur nature relationnelle et leur flexibilité (Kitchin, 2014). L'existence de ces très grands corpus de données, en provenance des plateformes numériques, explique pourquoi les

---

31. Nous ne nous risquerons pas, dans notre argumentation, à aborder la délicate question de la distinction entre le modèle et la simulation. Pour ceux qui souhaiteraient approfondir cela, nous ne saurions que trop suggérer la lecture des recherches en épistémologie de Franck Varenne (2007). Aux fins de notre démonstration, nous n'utiliserons que la notion de « modèle ».

32. « Le principe du rasoir d'Ockham consiste à ne pas utiliser de nouvelles hypothèses tant que celles déjà énoncées suffisent, à utiliser autant que possible les hypothèses déjà faites, avant d'en introduire de nouvelles [...] ». Source : [https://fr.wikipedia.org/wiki/Rasoir\\_d\\_%27Ockham](https://fr.wikipedia.org/wiki/Rasoir_d_%27Ockham)

chercheurs et experts de notre panel s'intéressent notamment au développement, par les méthodes computationnelles, de nouveaux modèles d'analyse de données. C'est en effet à l'édification de modèles adaptés à l'analyse de très grands corpus de données que paraît inviter l'émergence du phénomène des *big data*. Or, un modèle d'analyse de données vise, par définition, à « repérer les corrélations, à résumer et à cribler l'information empirique » (Varenne, 2010, p. 24). Les modèles permettent, parmi d'autres fonctions spécifiques, de « faciliter la compression et la synthèse de données disparates pour l'utilisation ultérieure » ou de « faciliter la reproduction et l'extrapolation d'une évolution observable » (Varenne, 2016, p. 14)<sup>33</sup>. D'après cette définition générale, la recherche de *corrélations* se situe donc au principe de la construction, par les méthodes computationnelles, de modèles d'analyse de données<sup>34</sup>. Cette recherche de corrélations qui était déjà au cœur du modèle statistique traditionnel se colore de nouvelles nuances avec l'approche computationnelle de très grands corpus. Les catégories classiques de la sociologie (âge, sexe, scolarité, profession, revenu) ne sont plus introduites *a priori* dans l'analyse. L'expert en mégadonnées (*data scientist*) oriente plutôt sa démarche inductive vers des corrélations inattendues qui émergent du corpus.

---

33. Bien que les modèles pour l'analyse de données ne visent pas à remplir une multitude de fonctions simultanément (en raison de contraintes méthodologiques), il est possible de recenser au moins neuf fonctions spécifiques pour lesquelles ceux-ci peuvent être développés (Varenne, 2016, p. 13-14). Les deux mentionnées ci-dessus – « la compression et la synthèse de données », mais aussi « la reproduction et l'extrapolation d'une évolution observable » – nous paraissent pertinentes pour notre argumentation, dans la mesure où elles semblent épouser les particularités des modèles d'analyse de données fondés sur les *big data*.

34. Bien que cette distinction intellectuelle entre les modèles d'analyse de données, de prédiction et d'explication présente l'avantage d'introduire la notion de « modèle », cette tripartition s'avère bien moins univoque dans la réalité. Pensons notamment à l'exploitation des *big data* par le secteur marchand. Lorsque ce dernier mobilise des méthodes computationnelles pour tirer profit de ces jeux de données, l'identification des corrélations avec les modèles d'analyse des données sert d'abord et avant tout à des fins prédictives (permettant ultimement aux acteurs d'agir et consommer).

## 2.5 L'hybridation des méthodes qualitatives et computationnelles

La majorité de nos enquêtés ont discuté non pas de la mixité de la statistique et du qualitatif, mais plutôt de la pertinence d'articuler les méthodes computationnelles aux approches qualitatives pour exploiter, dans les recherches sur les manières de s'informer, les données des plateformes en ligne. Pour nos interlocuteurs, l'hybridation de ces méthodes autoriserait l'étude des pratiques informationnelles d'un (très) grand nombre d'individus, sans toutefois renoncer au projet d'analyser les significations que les acteurs attribuent à leurs propres activités. Ce faisant, l'hybridation du computationnel et du qualitatif offrirait aux chercheurs la possibilité de *décrire* et *mesurer* les modalités par lesquelles un ensemble d'individus s'informent (et ce, à partir des corpus de données produites par les plateformes numériques), mais aussi d'en *comprendre* les significations en construisant des moments ou des espaces favorisant la parole des individus. Sans même expliciter (pour l'instant) ce qui distingue la description et la mesure, d'un côté, et la compréhension, de l'autre, force est de constater dès maintenant l'intention – explicite ou implicite – de conjuguer plusieurs modes de connaissance en procédant à l'hybridation de ces méthodes<sup>35</sup>. Ce constat rejoint d'ailleurs un récent état de l'art des méthodes pour appréhender les usages numériques qui met en relief « la nécessité d'articuler des méthodologies complémentaires, de multiplier les points de vue pour arriver à une compréhension fine, de l'intérieur, des pratiques. » (Beaudouin et Denis, 2014, p. 12).

Ces aspirations de nos enquêtés invitent à prendre conscience d'un double écueil – maintes fois constaté dans la littérature méthodologique en sciences sociales (voir par exemple, Beaudouin et Denis, 2014, p. 6) – dans la réalisation des recherches en sciences sociales sur les manières de s'informer. Les études qualitatives apportent une intelligibilité des pratiques informationnelles, mais leurs conclusions apparaissent difficilement généralisables puisqu'elles s'appuient sur des échantillons de

---

35. L'argument selon lequel l'hybridation des méthodes computationnelles et des approches qualitatives autoriserait la production de connaissance articulant la micro-analyse et la macro-analyse était également présent dans nos entretiens, quoique beaucoup plus rarement.

taille restreinte. Ce qui est une difficulté classique des approches qualitatives qui visent la compréhension fine et singulière des pratiques plutôt que la production de connaissances plus limitées, mais généralisables. Par ailleurs, les travaux reposant exclusivement sur des méthodes computationnelles produisent, certes, des données mettant en lumière des corrélations sous différentes formes (bilans chiffrés, graphiques, nuages de mots), mais on se prive alors de la possibilité d'en apprécier la signification profonde pour les usagers concernés. Les approches hybrides traduisent ainsi une volonté de dépasser les limites respectives de ces deux approches contrastées.

Nos entretiens comportent plusieurs éléments donnant un aperçu des modalités concrètes de cette hybridation. Ainsi, plusieurs chercheurs estiment nécessaire d'accorder le primat aux méthodes qualitatives. Cela reviendrait de cette manière à subordonner la mobilisation des méthodes computationnelles à l'approche visant ultimement à révéler les significations que confèrent les acteurs à leurs manières de s'informer. Dans cette perspective, les enquêtes qualitatives seraient ce par quoi les chercheurs définiraient au préalable leur problématique, construiraient leurs grilles d'analyse ou collecteraient d'autres informations *ad hoc*, en suivant les principes d'une « théorisation ancrée<sup>36</sup> », pour interpréter les données résultant des méthodes computationnelles. Ceci étant dit, l'articulation entre les deux types d'approches ne se résume pas à une relation de subordination. Grâce à l'intelligibilité qu'elles apportent à de volumineux corpus de données, les méthodes computationnelles pourraient à leur tour guider la réalisation d'une démarche qualitative approfondie, en aidant les chercheurs à identifier des sous-échantillons de personnes à interroger par la suite, ou en mettant à disposition des matériaux pertinents pour d'éventuels entretiens d'autoconfrontation.

---

36. « La théorisation ancrée n'est pas une méthode de recherche qualitative. C'est une méthode d'analyse de données "visant à générer inductivement une théorisation au sujet d'un phénomène culturel, social ou psychologique, en procédant à la conceptualisation et la mise en relation progressives et valides de données empiriques qualitatives" » (Paillé, 1996, p. 184). Voir : [https://fr.wikipedia.org/wiki/Th%C3%A9orisation\\_ancr%C3%A9e](https://fr.wikipedia.org/wiki/Th%C3%A9orisation_ancr%C3%A9e)



### 3. Quelques considérations épistémologiques

Pour répondre aux spécificités des processus de complexification des pratiques informationnelles, la communauté scientifique est amenée à procéder à des « expérimentations méthodologiques » qui passent par un questionnement au sujet des axes analytiques de la recherche sur les manières de s'informer, par une appropriation de différentes sources de données, mais aussi, comme nous l'avons vu, par un effort d'hybridation des méthodes qualitatives et computationnelles. Dans nos entretiens, ces ajustements sont à l'origine de réflexions de nature épistémologique. Nous souhaitons, en l'occurrence, prêter attention à trois interrogations récurrentes dans les discours de nos enquêtés : les modes de connaissance sous-jacents aux stratégies méthodologiques, les conséquences d'un possible « tournant computationnel » dans l'histoire des sciences sociales et la possibilité de mener une discussion a-disciplinaire sur les méthodes de recherche.

#### 3.1 Quelle intelligibilité peut-on espérer de ces choix méthodologiques ?

Nous voulons maintenant revenir sur une distinction jusqu'ici esquissée entre plusieurs modes de connaissance : la *description*, la *mesure* et la *compréhension*. Ces notions, fondamentales en épistémologie et en sciences sociales, offrent la possibilité de déterminer l'intelligibilité qu'apportent les recherches scientifiques – et leurs stratégies méthodologiques – à leurs objets<sup>37</sup>. Ainsi, la notion de « description » se révèle féconde pour saisir ce qui se joue avec le développement des modèles d'analyse de données associés aux méthodes computationnelles. Décrire consiste, au sens général du terme, « à s'en tenir à dire ce qui se montre ; c'est restituer fidèlement, sans les dénaturer, les phénomènes observés ; c'est indiquer comment les événements s'enchaînent effectivement les uns aux autres » (Soler, 2009, p. 78). À la lumière de cette acception, la description peut présupposer – mais ne

---

37. Cette sous-section s'appuie sur les distinctions proposées par Léna Soler dans son ouvrage *Introduction à l'épistémologie* (2009). Ceci étant dit, notre argumentation procède en simplifiant les propos de cette auteure et ne rend pas nécessairement justice aux nuances de son raisonnement.

présuppose pas nécessairement – une conception *réaliste* du savoir, d'après laquelle ce mode de connaissance permettrait, en reflétant avec exactitude des phénomènes empiriques, d'appréhender la réalité elle-même.

Ce présupposé mérite d'être souligné, en raison de l'attrait qu'il confère aujourd'hui aux méthodes computationnelles dans le contexte de l'émergence de « très grands corpus de données » issus du Web, corpus que l'on identifie, tant dans les médias qu'au sein des communautés de chercheurs, par l'expression *big data*. L'existence de corpus gigantesques de données construites par les plateformes numériques laisse entrevoir en effet la possibilité d'effectuer des *mesures* suffisamment riches et sophistiquées, à partir de modèles spécifiques d'analyse de données, pour développer des connaissances dont le contenu correspondrait, selon les tenants de ces nouvelles « sciences des données » (*data science, data-driven science*), à la « réalité même » des activités en ligne des usagers. Autrement dit, en fonction de postulats fortement positivistes, les chercheurs s'adonnant à l'analyse des très grands corpus de données (*big data*) prétendent que les mesures découlant des méthodes computationnelles seraient assimilables à des *instruments* permettant de décrire, sans introduire de biais, ce que feraient « réellement » les individus en ligne (et parfois même hors-ligne). Cette nouvelle posture positiviste – qui semble séduire aujourd'hui un sous-ensemble de chercheurs issus en particulier du monde des sciences naturelles (informatique, mathématiques, sciences de l'ingénieur, CSCW: *computer-supported cooperative work*) et peu conscients de la nature spécifique des données propres au monde des sciences humaines et sociales (SHS) – serait le véhicule de ce que nous pourrions appeler une « idéologie des *big data* » prétendant que l'émergence de ce nouveau type de mesures et de données issues directement du Web, de même

que l'invention des nouvelles méthodes computationnelles<sup>38</sup> qui y sont associées, permettrait de faire l'économie de tout cadre théorique pour traiter les données numériques (Anderson, 2008).

Même si ce type d'arguments et de provocations peut s'avérer choquant pour la plupart des praticiens en sciences humaines et sociales, il ne s'agit pas, pour employer une métaphore populaire, de « jeter le bébé avec l'eau du bain ». Il demeure imaginable, en effet, et sans souscrire aux présupposés positivistes des *data scientists* (« experts en mégadonnées »), de recourir aux méthodes computationnelles non seulement pour mesurer les activités en ligne des usagers, mais aussi pour appuyer l'invention d'un nouveau regard sociologique (Boullier, 2015a; 2015b). C'est une voie tracée notamment par le chercheur Richard Rogers et ses équipes de l'Université d'Amsterdam (Rogers, 2013). Ce dernier propose sous l'appellation de « *digital methods* », une approche sociologique nouvelle qui s'ancre dans l'analyse de données « natives » (c'est-à-dire produites directement par les interfaces du Web), recueillies sur et à partir du Web (« *methods of the medium* »). Cette nouvelle approche mobilise ces données « natives » non pas simplement pour décrire les usages des médias sociaux révélés par ces ensembles de données numérisées, mais plutôt pour équiper de nouvelle manière le regard des sciences sociales: « How can we study social media to learn something about society rather than about social media use? » (Rogers). Nous reviendrons sur ces approches dites des « *digital methods* » aux chapitres III et IV.

Dans nos entretiens, les mesures effectuées à partir des modèles d'analyse de données étaient envisagées non pas comme

---

38. Indiquons d'emblée que cette posture réaliste soulève de sérieuses critiques dans nos entretiens. Si nos enquêtés reconnaissent la pertinence – relative – des mesures élaborées à partir des modèles d'analyse de données des méthodes computationnelles, la plupart tendent à rallier la thèse selon laquelle ces instruments seraient davantage « des théories matérialisées » (Bachelard, 1995, p. 16). En d'autres termes, la validité de ces mesures ne dépendrait pas de leur degré de correspondance à la réalité elle-même, mais ferait intervenir d'autres critères scientifiques. Nous reviendrons sur cette question au chapitre III, à l'occasion d'une analyse des multiples choix que doivent effectuer les chercheurs en sciences sociales lorsqu'ils entendent exploiter des méthodes computationnelles.

un moyen de découvrir la réalité en soi des pratiques informationnelles, mais comme un moyen de les dénombrer à partir de conventions sociales déterminées. Pour expliciter cette conception non réaliste de la mesure, Alain Desrosières a été cité à plusieurs reprises par les chercheurs et experts de notre panel. Pour ce sociologue et historien de la statistique, la *mesure* s'apparente à un moment de la *quantification*, c'est-à-dire d'un processus impliquant « une série de conventions préalables, de négociations, de compromis, de traductions, d'inscriptions, de codages et de calculs conduisant à la mise en nombre » (Desrosières et Kott, 2005, p. 2). La mesure ne saurait par conséquent être identifiée à une simple évaluation numérique d'une grandeur existant dans la réalité. C'est précisément en ce sens que nos enquêtés estiment légitime de recourir à cet instrument pour appréhender les pratiques informationnelles actuelles.

Pour les personnes que nous avons interrogées, les méthodes qualitatives restent incontournables pour *comprendre* les manières par lesquelles les individus s'informent aujourd'hui. Or, d'après l'acception générale du terme, « [dans] le cas des faits humains, comprendre veut traditionnellement dire revivre jusqu'à un certain point, de l'intérieur et de manière intelligible, l'expérience d'autrui – être en mesure d'en reconstituer le sens et les motifs » (Soler, 2009, p. 84). La compréhension suppose, partant de là, de mettre en lumière les facteurs subjectifs des comportements humains : les sentiments, les désirs, les valeurs morales, les convictions politiques, les intentions, les croyances religieuses ou, pour prendre un autre exemple, les intérêts égoïstes.

En deçà de cette acception se trouve l'opposition classique, en épistémologie et en sociologie, entre la *compréhension* et l'*explication*. Bien qu'il soit impossible, dans les limites de ce rapport, d'explorer toutes les subtilités de cette dichotomie, cette dernière plonge ses racines dans la philosophie de Wilhelm Dilthey, laquelle est d'ailleurs à l'origine d'une formule célèbre : « nous expliquons la nature, nous comprenons la vie psychique »

(1992)<sup>39</sup>. Cette formulation énoncée au 19<sup>e</sup> siècle pourrait toutefois être considérée aujourd'hui comme relativement dépassée dans la mesure où Max Weber insistera ensuite sur la complémentarité, plutôt que sur une opposition radicale, entre la compréhension et l'explication<sup>40</sup>. Pour Wilhelm Dilthey, la compréhension se définit d'abord par opposition à l'explication, l'une conduisant les chercheurs en sciences humaines et sociales à se mettre à la place de leurs semblables pour appréhender, voire interpréter, les significations de leurs pratiques et l'autre étant le propre des raisonnements nomologiques des sciences de la nature. Nos enquêtés développent en l'occurrence des analyses similaires qui ne concernent pas la distinction entre sciences humaines et sociales (SHS) et sciences de la nature<sup>41</sup>, mais plutôt les différences entre méthodes qualitatives et computationnelles. Les méthodes qualitatives sont clairement identifiées, dans nos entretiens, à ce par quoi il est possible d'accéder aux significations que donnent les individus à leurs pratiques informationnelles. En travaillant à l'hybridation des méthodes, les chercheurs confèrent donc aux approches qualitatives la fonction d'apporter une *compréhension* aux manières de s'informer des acteurs, cette compréhension passant autant par la mobilisation de techniques d'enquête qualitatives (dans le but de recueillir et de traiter des données qualitatives) que par l'interprétation des mesures quantitatives des modèles d'analyse de données.

Pour revenir brièvement sur la distinction entre explication et compréhension du point de vue des sciences humaines et sociales (SHS), il est certain qu'en se constituant, au 19<sup>e</sup> siècle,

---

39. La philosophie de Wilhem Dilthey n'a pas été citée dans nos entretiens, mais l'opposition que cet auteur énonce entre la compréhension et l'explication s'avère éclairante pour saisir comment notre panel d'experts et de chercheurs envisagent certains aspects des méthodes qualitatives.

40. « [...] c'est là encore un des points par lesquels Weber prend distance vis-à-vis d'une conception trop autonome de la méthode compréhensive, l'explication n'est point séparable de la compréhension. Aucune interprétation, si évidente soit-elle, ne peut être considérée comme valable indépendamment de l'application à son égard des méthodes de l'implication causale. Mais réciproquement, les explications causales ne sont réellement satisfaisantes que lorsqu'elles sont réellement interprétées de manière significative. » (Hirschhorn, 1999, p. 99)

41. Le syntagme « sciences de l'esprit » utilisé par Dilthey renvoie, au 19<sup>e</sup> siècle, à ce que nous appelons aujourd'hui les sciences humaines et sociales.

la sociologie était portée par « l'ambition générale d'expliquer les phénomènes sociaux et historiques » (Ansart, 1999, p. 212). Le principe d'une autonomie de l'explication sociologique sera posé de manière fondatrice par Émile Durkheim qui sera suivi plus tard notamment par les tenants des approches structurales et par l'école de Pierre Bourdieu. C'est Max Weber qui s'opposera à Durkheim en faisant de la compréhension « l'une des méthodes indispensables de la sociologie, mais (qui) n'invalide pas la recherche des explications » (Ansart, 1999, p. 212). Cette opposition entre perspectives plutôt structuralistes et approches plutôt compréhensives marque encore aujourd'hui la dynamique de développement du champ des sciences humaines et sociales. Dans la mesure où la majorité des chercheurs interrogés s'adonnaient en particulier aux approches qualitatives, il n'est pas surprenant de constater qu'une grande partie d'entre eux se sentaient plus proches d'une sociologie compréhensive de l'action, laquelle se situe dans le prolongement des travaux de type ethnographique de l'École de Chicago (Chapoulie, 2001) – et notamment d'Howard S. Becker (2002, 2004) – qui emprunte particulièrement aux postulats épistémologiques d'une phénoménologie sociale (Blin, 2010).

### **3.2 Peut-on parler « d'un tournant computationnel » en sciences sociales ?**

Une autre question épistémologique s'avère prégnante dans nos entretiens. Si elle n'est pas toujours explicitement discutée par nos enquêtés, elle n'en constitue pas moins la toile de fond de leur réflexion sur l'actualité des méthodes en sciences sociales. Il s'agit, pour le dire schématiquement, d'une interrogation sur la possibilité d'une rupture dans l'évolution historique des sciences sociales, engendrée par l'apparition de nouveaux

dispositifs méthodologiques<sup>42</sup>. Motivée par l'hypothèse d'un « tournant computationnel », c'est-à-dire l'idée que nous traverserions aujourd'hui une conjoncture singulière, marquée par la rencontre de la généralisation des méthodes computationnelles et de la disponibilité des *big data*, cette question a suscité des réactions à la fois contrastées et nuancées<sup>43</sup>.

Nos entretiens comportent plusieurs arguments à ce sujet, qu'il semble possible de regrouper en trois grandes positions. La première consiste à voir dans ce « tournant computationnel » un moment de discontinuité dans l'histoire des sciences sociales. Selon cette logique, la configuration sociotechnique actuelle obligerait les sciences sociales non pas à procéder à de simples ajustements, mais à adopter de nouvelles conventions par rapport aux générations précédentes des sciences sociales. Ce faisant, nous assisterions à l'émergence d'une troisième génération de sciences sociales, connaissant de profondes transformations touchant autant leurs appareils conceptuels, leurs problématiques, leurs objets que leurs dispositifs de collecte et de traitement des données (Boullier, 2015a).

Moins radicale, mais aussi plus répandue parmi les personnes que nous avons interrogées, la deuxième position demeure plus prudente sur les conséquences épistémologiques de cette rencontre des méthodes computationnelles et des *big data*. Elle prend acte de la naissance des nouvelles méthodes numériques,

---

42. Le texte de Mike Savage et Roger Burrows sur « The Coming Crisis of Empirical Sociology » publié en 2007, quoiqu'ancré dans une sociologie britannique en crise, a suscité des discussions considérables à propos de l'émergence des « données transactionnelles », expression qui préfigurait l'idée de *big data*. Leur texte ultérieur de 2014 fait état du cheminement de leur questionnement sur les enjeux des *big data* pour le développement des méthodes en sciences sociales. Pour de nombreux sociologues, ces textes ont offert un cadrage davantage déterminant que ceux en « digital methods » ou en « computational social sciences ». Merci à Éric Dagiral de nous avoir fait cette remarque.

43. C'est dans un article publié en 2009, intitulé « Life in the network: the coming age of computational social science », que se trouve pour la première fois formulée explicitement l'hypothèse de la naissance d'un tournant computationnel en sciences sociales (Lazer *et al.*, 2009). Les auteurs font le diagnostic d'une rupture dans l'histoire des sciences sociales en raison de la configuration sociotechnique actuelle et appellent la communauté scientifique à embrasser les potentialités de ces transformations dans le développement de connaissances scientifiques sur les réalités individuelles ou sociales (p. 2).

en évitant toutefois de se prononcer sur l'éventualité d'une disparition des sciences sociales ou, tout du moins, d'une dissolution des sciences sociales telles qu'elles se sont instituées depuis plus d'un siècle. Elle s'attache davantage à interroger les possibilités qu'offrent ces méthodes numériques aux chercheurs en sciences sociales et à identifier les limites inhérentes à ces dernières. Du reste, cette analyse des potentialités des méthodes computationnelles dans le contexte des *big data* peut être effectuée autant dans un cadre disciplinaire que dans un contexte interdisciplinaire. Cette deuxième position rassemble, par exemple, des réflexions sur les modalités concrètes de la contribution de ces méthodes numériques à la construction de connaissances en sciences sociales, sur les manières d'articuler les méthodes numériques aux méthodes conventionnelles ou encore sur l'intégration des méthodes numériques à une démarche compréhensive. Plus « pragmatique » que les deux autres, en un certain sens, cette position convient implicitement de l'existence d'un « tournant computationnel », mais en s'abstenant de tirer des conclusions sur les conditions de validité des connaissances en sciences sociales aujourd'hui.

La troisième position témoigne quant à elle d'un point de vue sociologique sur cette question. Elle présente l'avantage de formuler plusieurs observations, ayant parfois valeur de mise en garde, par rapport à cette discussion sur une possible rupture dans l'histoire des sciences sociales. Ces observations conduisent finalement à relativiser la pertinence de ce débat, d'une part, et à défendre l'existence d'une continuité dans les recherches en sociologie, d'autre part. Au préalable, cette position reconnaît sans difficulté l'intérêt que peuvent avoir les innovations méthodologiques sur le plan des outils de collecte et de traitement des données pour la recherche en sciences sociales, ainsi que les mérites des jeux de données provenant des plateformes en ligne (du fait de leur granularité et de leur volume). Les tenants de cette position se refusent néanmoins à déduire de ces évolutions méthodologiques une révolution épistémologique. Les normes du travail sociologique ne connaîtraient en effet pas d'altérations majeures aujourd'hui : il s'agirait toujours d'adopter – entre autres – un point de vue critique sur les données en analysant leur champ d'application légitime, leurs biais ou leurs limites ; d'élaborer avec réflexivité des dispositifs efficaces de collecte et



de traitement de données; de s'assurer de l'adéquation entre les données recueillies et la problématique de recherche; de proposer une interprétation à la fois fine et appuyée de ces données. Quand bien même les études sociologiques actuelles auraient recours aux données des plateformes en ligne ou à de nouveaux outils méthodologiques, elles n'en demeureraient pas moins, dans cette perspective, conventionnelles au regard des normes du travail sociologique.

### 3.3 Une discussion méthodologique sans ancrage disciplinaire ?

Un autre problème épistémologique a été abordé à plusieurs reprises dans nos entretiens, souvent d'ailleurs par les mêmes personnes qui argumentaient en faveur de la troisième position énoncée précédemment. Cette interrogation porte en l'occurrence sur le bien-fondé (ou non) d'une discussion *a-disciplinaire* sur la place des méthodes numériques dans les recherches en sciences sociales. Dans la mesure où la notion de « sciences sociales » désigne un large spectre de connaissances scientifiques, appartenant à des champs disciplinaires multiples et variables (la psychologie, la science politique, l'anthropologie, la linguistique, la sociologie, etc.), la légitimité d'une telle discussion ne saurait être tenue pour acquise. Cette dernière repose en effet implicitement sur la possibilité d'une *autonomisation* des questions méthodologiques par rapport aux disciplines en sciences sociales. Or, ce présupposé, à l'œuvre dans une partie de la littérature scientifique autour des « Digital Methods » (Rogers, 2009, 2013) ou des « Critical Digital Studies » (Kroker et Kroker, 2008) fait doublement problème. Il contribue à évacuer subrepticement les problématiques propres à chaque champ de recherche en sciences sociales et, *a contrario*, à « invisibiliser » les problématiques de recherche dont les dispositifs méthodologiques – quels qu'ils soient – sont les expressions concrètes. Tout se passe comme si, dans cette perspective, les stratégies méthodologiques et les techniques d'enquête pouvaient être appréhendées sans référence aux problématiques respectives à chaque discipline en sciences sociales et, plus généralement, à des questions de recherche. Autrement dit, c'est l'articulation entre la méthode et la définition d'une problématique qui se perd avec cette autonomisation.

Les auteurs de *Sociologie d'Internet* (Beuscart, Dagiral et Parasio, 2016) exposent dans leur introduction une argumentation permettant de saisir, à partir de l'exemple de la sociologie, ce qui se joue avec ces discussions a-disciplinaires sur les méthodes numériques en sciences sociales. Pour ces auteurs, les études sociologiques d'Internet s'intéressent ultimement à trois questions transversales, à savoir: ce que fait internet à « la capacité d'action des individus », « aux relations entre individus » (du point de vue de l'égalité entre eux) et « à la visibilité de la société à elle-même » (Beuscart, Dagiral et Parasio, 2016, p. 22). Or, loin d'apparaître avec internet, cette triple interrogation appartient en réalité à l'histoire de la sociologie et demeure constitutive de cette discipline. Dès lors, l'autonomisation des discussions sur les méthodes numériques, en déconnectant les aspects méthodologiques des ancrages disciplinaires, risque de reléguer silencieusement, sans réflexion préalable, ces trois problématiques sociologiques essentielles.

\* \* \*

Développée à partir de 24 entretiens réalisés avec des chercheurs francophones en sociologie, en communication et en gestion, l'argumentation précédente dresse le portrait des pratiques informationnelles actuelles. Elle révèle la complexité grandissante de ces dernières, en raison d'une diversification des sources d'information, de la fragmentation de la réception, de la généralisation du *multitasking*, de la multiplication des supports, mais aussi de l'instabilité des pratiques informationnelles. Ce contexte incite les chercheurs à expérimenter de nouvelles techniques de collecte et d'analyse de données pour appréhender les manières de s'informer contemporaines. Il invite, autrement dit, au « bricolage méthodologique ». Il requiert enfin de penser à nouveaux frais un certain nombre de questions épistémologiques. Ayant présenté les enjeux et les problèmes entourant les méthodes de recherche en lien avec le numérique, nous précisons dans le prochain chapitre notre réflexion en distinguant les grandes catégories de méthodes permettant d'étudier les pratiques informationnelles actuelles.

# 3

## Les méthodes de recherche en sciences sociales en lien avec le numérique

**D**ans ce chapitre, nous souhaitons exposer et détailler les stratégies méthodologiques dont s'emparent les chercheurs en sciences sociales pour étudier les pratiques informationnelles actuelles. À cet effet, nous mobiliserons le contenu de nos 24 entretiens, mais aussi, plus largement, la littérature scientifique offrant des réflexions sur les modalités mêmes de l'extraction et de l'analyse de données sur les manières de s'informer. Pour éviter d'inutiles malentendus, précisons d'emblée que ce chapitre ne prétend pas dresser un tableau complet des dites méthodes. Il s'agit plutôt de s'arrêter sur les éléments revenant avec récurrence dans nos entretiens, ainsi que dans la littérature

scientifique (francophone et anglophone)<sup>44</sup>. Quoique partielle, cette recension entend appréhender ces méthodes à la lumière de leurs possibles articulations avec des problématiques déterminées. Loin d'être de simples outils, que nous pourrions comparer sur des critères strictement instrumentaux (leur coût, leur vélocité, leur facilité d'utilisation, etc.), ces méthodes n'ont de pertinence que dans leur adéquation avec certaines questions de recherche. Nous éviterons ainsi de donner à ce chapitre la forme d'un catalogue, négligeant les articulations entre les moments de la problématisation – celle-ci étant liée à une théorisation implicite ou explicite – et la conception d'un dispositif méthodologique au cours du processus de recherche.

En empruntant à Max Weber son concept d'*idéal-type* (Weber, 1965)<sup>45</sup>, nous amorcerons notre argumentation en construisant une typologie des principales approches méthodologiques dans les recherches sur les pratiques informationnelles actuelles. Nous procéderons alors en identifiant les caractéristiques saillantes de ces perspectives méthodologiques, de manière à en déterminer les spécificités et à rendre possibles des comparaisons entre elles. Bien qu'elle ait été initialement élaborée dans le contexte des études sur la science et la technologie (*Science and Technology*

---

44. Pour étudier la littérature scientifique, nous avons procédé en demandant à chaque personne interrogée des recommandations de lecture, tout en consultant divers ouvrages de référence concernant les méthodes computationnelles en sciences sociales (*handbooks*, manuels, *companions*, etc.). Cette stratégie présentait le double avantage de nous aider à préciser le positionnement de nos enquêtés par rapport aux lectures qu'ils pouvaient nous suggérer, d'une part, et à commencer à entrevoir les contours des méthodes exploitables dans le cadre de recherches en sciences sociales sur les pratiques informationnelles, d'autre part. À partir des résultats de cette première étape, nous avons effectué un travail de repérage des références bibliographiques les plus fréquemment citées par la communauté scientifique, sans toutefois exclure les « heureuses découvertes » de la sérendipité. Il va sans dire qu'une telle stratégie, parce qu'elle n'offre pas la possibilité d'élaborer une revue de la littérature scientifique systématique et exhaustive, comporte nécessairement des failles et des omissions. C'est là l'une des limites de notre démarche.

45. « L'idéal-type est un tableau de pensée, il n'est pas la réalité historique ni surtout la réalité "authentique", il sert encore moins de schéma dans lequel on pourrait ordonner la réalité à titre d'exemplaire. Il n'a d'autre signification que d'un concept limite purement idéal, auquel on mesure la réalité pour clarifier le contenu empirique de certains de ses éléments importants, et avec lequel on la compare. Ces concepts sont des images [...] » (Weber, 1965, p.143).

*Studies*), nous nous inspirerons d'une distinction, formulée par Noortje Marres, au sujet des stratégies méthodologiques à l'œuvre dans les recherches en sciences sociales en lien avec le numérique (Marres, 2012). Partant, nous différencierons cinq approches : les méthodes conventionnelles ; les ethnographies virtuelles (ou ethnographies « en ligne ») ; les méthodes computationnelles tirant profit des *big data* ; les méthodes numériques (*digital methods*) ; et finalement, les méthodes numériques quali-quantitatives<sup>46</sup>.

Nous aborderons, dans un second temps, les problèmes que peuvent engendrer les méthodes en lien avec le numérique. Nous nous arrêterons sur les enjeux organisationnels, politiques (au sens des rapports stratégiques ou de pouvoir), économiques et éthiques accompagnant potentiellement leur mise en œuvre. Les enjeux organisationnels relèvent pour l'essentiel des modalités de collaboration entre les chercheurs en sciences sociales et les informaticiens. Les enjeux politiques concernent les possibles tensions entre les intérêts des chercheurs et ceux des propriétaires des plateformes en ligne, les finalités des uns et des autres n'étant, à l'évidence, pas toujours harmonieuses. Les questions économiques ont trait aux ressources financières requises pour avoir accès aux données numériques. Les inégalités économiques engendrent en effet de sérieuses discriminations entre les structures de recherche, les plus précaires se voyant *de facto* interdire la possibilité de conduire des études reposant sur la collecte et l'analyse des traces des *big data*. Nous prêterons enfin attention aux enjeux éthiques en lien avec le caractère présumé « public » des données numériques.

## 1. Idéaux-types des stratégies méthodologiques

Les stratégies méthodologiques que nous entendons appréhender en tant qu'idéaux-types entretiennent toutes un lien avec le numérique. Ce lien comporte cependant plusieurs dimensions.

---

46. Au moment de la finalisation de ce rapport, un numéro spécial de la *Revue française de sociologie* (2018/3, vol. 59) intitulé « Big Data, sociétés et sciences sociales » vient tout juste de paraître. Nul doute que lors d'une future édition du présent rapport, la lecture de ce numéro contribuera à complexifier notre tableau synthétique.

Le numérique peut d'abord être l'objet – un ensemble de phénomènes – pour lequel ces méthodes sont mobilisées à des fins de recherche. Dans le cadre des pratiques informationnelles actuelles (dont les contours empiriques sont esquissés dans le chapitre II), ces méthodes permettent en l'occurrence de collecter et d'analyser des données relatives aux usages des médias numériques. Ce lien peut, en second lieu, prendre la forme d'une numérisation des stratégies méthodologiques, ce processus impliquant soit un travail sur des données de recherche préexistantes, soit l'usage des techniques d'enquête dites « conventionnelles », c'est-à-dire qui ne s'appuient pas sur une collecte de traces en ligne, (qui impliquent, par exemple, l'usage de logiciels facilitant l'analyse qualitative de contenus textuels, ou de tableurs pour manipuler des données quantitatives, etc.) ou les deux. Ce lien se manifeste enfin avec les méthodes matérialisées dans les plateformes en ligne (méthodes dites « nativement numériques »), ces méthodes embarquées étant précisément ce par quoi les plateformes produisent et traitent, selon des logiques déterminées, leurs données numériques (*web analytics*). Nous reviendrons, à l'occasion de notre présentation des méthodes numériques (*digital methods*), sur la distinction entre méthodes numérisées et méthodes dites « nativement numériques ».

En recourant aux recherches de Noortje Marres pour construire des idéaux-types des méthodes en lien avec le numérique, nous n'aurons d'autre choix que de soustraire son travail de la problématisation pour laquelle celui-ci a été effectué. En effet, les distinctions qu'elle opère entendent non pas servir un projet de cartographie des méthodes actuelles en lien avec le numérique, mais plutôt corroborer l'hypothèse d'une « redistribution des méthodes en sciences sociales » dans le contexte du numérique. Cette hypothèse consiste, pour faire vite, à concevoir la recherche scientifique comme un processus impliquant une variété d'acteurs : chercheurs, organismes subventionnaires, enquêtés, amateurs, concepteurs d'infrastructures, informaticiens, etc. C'est au prisme de la division du travail scientifique que Marres différencie quatre formes de méthodes en lien avec le numérique. Notre démarche s'inspire assez librement de son travail de catégorisation, que nous adaptions à notre projet d'identifier les perspectives méthodologiques discutées

aujourd'hui de manière récurrente dans les échanges entre les chercheurs en sciences sociales.

## 1.1 Les méthodes conventionnelles

Avec les méthodes conventionnelles, nous abordons des stratégies familières et relativement bien connues. Qu'elles soient quantitatives, qualitatives ou mixtes, elles partagent plusieurs caractéristiques communes. Elles jouissent, en premier lieu, d'une certaine stabilité. Bien qu'elles aient pu faire l'objet d'ajustements, de complexifications ou de corrections en fonction des débats agitant la communauté scientifique, il n'en reste pas moins qu'elles accompagnent le développement des sciences sociales depuis leur naissance. Les méthodes conventionnelles bénéficient, en second lieu, de la reconnaissance des chercheurs en sciences sociales. Elles suscitent des critiques qui, à bon droit, questionnent leurs potentialités heuristiques, leurs présupposés ou leurs insuffisances, mais qui remettent rarement en cause leur légitimité. En troisième lieu, les méthodes conventionnelles n'apparaissent pas affectées par le numérique. Identiques à elles-mêmes, elles s'avèrent imperméables aux remous provoqués par le numérique. Autrement dit, elles s'inscrivent pleinement dans la continuité de l'histoire des sciences sociales, et ce, quand bien même elles permettraient de collecter et d'analyser des données sur des phénomènes numériques. Non sans une pointe d'ironie, à laquelle nous ne souscrivons pas, Noortje Marres estime qu'il s'agit là d'une vision conservatrice des méthodes de recherche en sciences sociales, en somme, les *methods as usual* (2012, p. 149). Au fur et à mesure que nous réalisons nos entretiens avec les chercheurs, nous avons pu identifier un noyau significatif de chercheurs mobilisant les méthodes conventionnelles sans pour autant s'identifier à un courant conservateur en sciences sociales. En recourant systématiquement à ces méthodes dites conventionnelles, ils affirmaient au contraire leur pratique professionnelle assumée du métier de sociologue.

## 1.2 Les ethnographies virtuelles

L'émergence de l'ethnographie virtuelle – ethnographie également dite « en ligne » – doit beaucoup aux travaux pionniers de Christine Hine. Cette chercheuse jette les bases de cette

approche au tournant du 21<sup>e</sup> siècle, avec la publication de deux ouvrages, portant respectivement les titres *Virtual Ethnography* (2000) et *Virtual Methods* (2005). Loin d'être désuète dans le champ des *Internet Studies*, l'ethnographie virtuelle a été, au fil des discussions dans la communauté scientifique, à la fois précisée et complexifiée, ce dont témoigne par exemple l'ouvrage *Ethnography for the Internet: Embedded, Embodied and Everyday* (Hine, 2015). L'auteure se situe en continuité avec la tradition anthropologique ; elle poursuit notamment le travail de Clifford Geertz (1973) visant à produire une « description dense et détaillée » (*thick description*) de la réalité observée (Becker, 2004, p. 77-81). Hine constate que la plupart des expériences quotidiennes des individus sont aujourd'hui traversées par le recours à une variété de formes de « communication médiatisée par ordinateur » (ou plus généralement : médiatisée par l'informatique et le numérique) (*computer-mediated communication*). Aussi, l'observation ethnographique de ce type de situations suppose une adaptation des méthodes et techniques pour sonder en profondeur les manières par lesquelles les personnes donnent sens à leur existence dans un tel contexte. Trois propriétés dessinent les contours de cette perspective anthropologique : sa pluralité, son adaptabilité et sa nature qualitative.

Christine Hine insiste, dans ses plus récentes publications (2017), sur la diversification des formes de l'ethnographie virtuelle, ce processus étant parallèle aux évolutions que connaissent les médias numériques eux-mêmes. La pluralité des ethnographies virtuelles s'explique de son point de vue par l'impératif d'épouser la réalité – pour le moins fluctuante – des phénomènes numériques. La variété des démarches méthodologiques découle également de la multiplicité des choix théoriques, des problématiques et, le cas échéant, des points d'appui normatif privilégiés par les chercheurs dans leurs études des phénomènes numériques. Voilà pourquoi Christine Hine parle à présent non plus d'une ethnographie virtuelle singulière, mais plutôt d'une pluralité de formes d'ethnographies en ligne (Hine, 2017).

Au-delà de la diversification de ses formes, notons que cette approche a pour ambition d'adapter les méthodes conventionnelles de l'observation ethnographique au domaine numérique. Les ethnographies virtuelles reprennent par là même les principes de l'enquête de terrain, mais en affrontant les difficultés



épistémologiques, éthiques et logistiques de l’accomplissement d’une telle démarche dans le contexte particulier du numérique. Par ailleurs, les recherches en sciences sociales reposant sur cette perspective rencontrent aussi le problème de la définition des frontières de leur terrain, celui-ci étant délimitable selon différentes modalités : se focaliser sur des espaces exclusivement numériques, observer un groupe d’acteurs autant dans leurs interactions hors ligne qu’en ligne, suivre un flux de communication concernant une thématique déterminée, s’intéresser à l’interconnexion entre plusieurs terrains, etc.

Bien que les ethnographies virtuelles n’aient pas été développées à l’origine dans un rapport d’opposition contrastée aux méthodes computationnelles, et plus encore aux méthodes computationnelles recourant aux *big data*, force est de constater que celles-ci se positionnent aujourd’hui par rapport à celles-là dans la littérature scientifique récente (Hine, 2017; Boyd et Crawford, 2012). Tout se passe en effet comme si la légitimité de cette stratégie ethnographique requérait un effort de démarcation d’avec l’approche de type *big data*. Quoi qu’il en soit, ce rapport d’opposition présente pour nous l’avantage de souligner la nature qualitative des ethnographies virtuelles. Celles-ci trouvent leur raison d’être dans la conception d’études de portée restreinte, fondées sur des corpus limités de données. Elles apportent une intelligibilité aux phénomènes numériques, dans la mesure où elles permettent de cerner les significations de pratiques médiatisées par les technologies de communication numérique (en se donnant les moyens d’accéder à l’expérience vécue des usagers ou aux contextes de leurs pratiques sociales). Bref, ces approches ethnographiques permettent une *compréhension* des pratiques, au sens épistémologique du terme. Par ailleurs, cette compréhension peut également se nourrir d’une autoethnographie du chercheur (Markham, 1998; Dumitrica et Gaden, 2008), cette dernière manière de faire offrant l’occasion d’analyser les phénomènes numériques par le biais d’une posture réflexive interrogeant la propre expérience d’usager du chercheur, dans ses dimensions cognitive et affective.

### 1.3 Les méthodes computationnelles exploitant les big data (approche dite « big methods »)

Pour désigner les méthodes computationnelles se saisissant des potentialités des très grands corpus de données (*big data*), Noortje Marres parle de l'approche de type *big methods* (2012). Cette stratégie méthodologique consiste ultimement à tirer profit des grands corpus de données numériques que les plateformes en ligne produisent et traitent. Contrairement à une représentation assez répandue, à laquelle invitent les discours marketing et commerciaux autour des *big data*, ces données n'ont pas pour seule spécificité d'être volumineuses. Pour Dominique Boullier, les traces des *big data* – nous aurons l'occasion de revenir sur la signification de la notion de « trace<sup>47</sup> » – ont pour propriété d'être à la fois volumineuses, variées et véloces (2015b, p. 810). Au-delà de leur quantité, certes impressionnante, elles se caractérisent par l'extrême variabilité de leur format et de leur contenu, mais aussi par la rapidité et la réactivité de leur production par les plateformes en ligne. À partir d'une revue de littérature sur cette question, Rob Kitchin définit quant à lui les données de type *big data* du point de vue de leur volume, de leur vélocité, de leur exhaustivité, de la finesse de leur granularité, de leur nature relationnelle et de leur flexibilité (2013, p. 262)<sup>48</sup>. Saisir les particularités de l'approche dite *big methods* suppose par là même de

---

47. Voir la définition de « trace » dans le glossaire que nous fournissons à la fin de ce rapport.

48. Bien que nous ne puissions pas entrer ici dans les détails de cette discussion, disons simplement que le critère de l'exhaustivité pour définir les *big data* fait l'objet de désaccords dans la communauté scientifique. Si Rob Kitchin estime qu'il s'agit bien d'une spécificité des *big data*, Dominique Boullier conçoit au contraire l'exhaustivité comme un principe de validation propre aux recherches sociologiques dites de « première génération ». Or, celles-ci voient le jour à la fin du 19<sup>e</sup> siècle avec la sociologie durkheimienne (Boullier, 2015a). L'un appréhende donc l'exhaustivité comme une spécificité des méthodes développées dans le contexte sociotechnique contemporain alors que l'autre l'attribue à une approche cognitive existant depuis plus d'un siècle.

se défaire d'une conception trop simplificatrice des données de type *big data*<sup>49</sup>.

Dans ses réflexions sur les *big data*, Dominique Boullier distingue la notion de « trace » de celle de « donnée » (2015b, p. 809-810). Cette distinction s'avère pertinente pour saisir ce qui caractérise l'approche de type *big methods*. En effet, les traces ne doivent pas être confondues avec les données numériques qui sont formatées et structurées de manière à fournir des informations sur les attributs sociaux, économiques, démographiques ou politiques d'un ensemble d'individus. Si une donnée numérique peut être techniquement manipulée en tant que trace dans le cadre des *big data*, les traces ne se réduisent pas aux données des systèmes d'information permettant, par exemple, la gestion de clients, d'administrés ou d'électeurs. La notion de « trace » recèle ainsi une extension sémantique de bien loin supérieure à celle de « donnée » : « [les traces] peuvent aller de signaux ('bruts', générés par des objets) à des verbatim non structurés qui se propagent sous la forme de mèmes (ou de citations), elles peuvent être des métadonnées (plus que les contenus d'un tweet, ces métadonnées sont très riches et aisément calculables), des traces (liens, clics, like, cookie) exploitées en bases de données par les

---

49. Pour l'instant, les traces des *big data* manipulées par les sociologues ne semblent pas aussi « massives » que dans les recherches appartenant au domaine des sciences de la donnée (*data science*). En disant cela, nous n'entendons pas suggérer l'impossibilité de recourir à des corpus volumineux de données numériques dans des études en sciences sociales, mais plutôt donner un aperçu plus en adéquation avec leur réalité actuelle. Gilles Bastin et Paola Tubaro expliquent d'ailleurs à ce sujet : « Malgré leur originalité et ces changements méthodologiques majeurs en matière d'analyse, les données utilisées par nos auteurs ne sont pas toujours massives au sens strict du terme. La taille des données numériques utilisées dans notre discipline ne dépasse pas encore de façon significative celle des grandes bases de données de la statistique publique classique, comme le recensement ou l'enquête « Emploi », et le temps ne semble pas encore arrivé où les sociologues devront recourir à des outils nouveaux en matière de traitement des données comme le calcul parallèle qui occupe une partie des collègues qui se reconnaissent dans d'autres disciplines comme des praticiens des *big data*. » (Bastin et Paola, 2018, p. 384)

opérateurs de plateformes (GAFAM) » (Boullier, 2015b, p. 809)<sup>50</sup>. Au sens strict, il serait par conséquent préférable de parler non pas de données, mais de traces des *big data*, même si cette distinction n'est pas toujours formulée explicitement dans les publications scientifiques. Nous devinons, à la lumière de cette acception, que la notion de « trace » recouvre des réalités empiriques très diversifiées dans la mesure où ces traces sont engendrées dans des situations sociales – d'énonciation ou d'action – non moins variées<sup>51</sup>. Par ailleurs, il nous apparaît nécessaire de distinguer minimalement les traces générées de manière involontaire ou non consciente par les usagers, faisant l'objet d'un traitement par les algorithmes des plateformes en ligne (données de transaction, de navigation, etc.), d'une part, et les traces déposées de façon délibérée par les internautes (commentaires sur Facebook, etc.), d'autre part<sup>52</sup>.

Bien que les potentialités de l'exploitation des traces des *big data* pour la recherche en sciences sociales fassent toujours l'objet de légitimes discussions, sinon de controverses (Boyd et Crawford, 2012), il faut bien constater qu'un tel débat n'aurait pas de pertinence sans la disponibilité de ressources computationnelles – matérielles et algorithmiques – en adéquation avec les exigences

---

50. Dominique Boullier distingue trois catégories de traces des *big data*. Il y aurait d'abord « ces traces faites de clics, de likes ou d'autres marqueurs d'attention éphémère et non verbale », puis « les commentaires qui permettent des traitements lexicaux se rapprochant de ce qui est fait dans les études d'opinion sans avoir aucunement le statut des méthodes établies de mesure de l'opinion » et, enfin, les liens hypertexte qui constituent des marques d'attachement préférentiel et qui peuvent alors être souvent exploités comme le sont les données des analyses de réseaux sociaux ou comme des marqueurs d'autorité tels que la scientométrie les analysait traditionnellement sous forme de citations » (2015b, p. 809).

51. Cette diversité des traces n'est pas sans conséquence pour les chercheurs en sciences sociales, dans la mesure où chaque type de traces dessine des potentialités et des limites du point de vue des possibilités analytiques.

52. Cette distinction révèle en outre une différence notable entre les *big methods* et l'ethnographie en ligne, dans la mesure où cette dernière conduit les chercheurs à se focaliser sur la collecte et l'analyse d'un nombre limité de traces d'activités intentionnelles, laissées par les usagers sur les plateformes en ligne. Les *big methods* entendent manipuler des corpus non seulement plus volumineux de traces, mais aussi beaucoup plus diversifiés (traces intentionnelles et non intentionnelles). Le rapport de recherche de Nicole Gallant, Guillaume Latzko-Toth et Madeleine Pastinelli (2015) intitulé *Circulation de l'information sur les médias sociaux pendant la grève étudiante de 2012 au Québec* nous paraît représentatif de ce trait des ethnographies en ligne.

de leur manipulation. De ce fait, ce que Noortje Marres appelle l'approche *big methods* se trouve en réalité à la confluence entre, d'une part, l'accessibilité de données aux propriétés particulières (les traces des *big data*) et, d'autre part, la possibilité de recourir à des technologies computationnelles suffisamment puissantes et efficaces pour traiter ces données. De plus, l'exploitation des *big data* s'accompagne souvent de l'utilisation d'outils de visualisation (graphes, cartographies, nuages de mots...) pour représenter les résultats du traitement algorithmique d'un ensemble déterminé de traces. La compréhension de l'approche *big methods* ne peut donc faire l'économie d'une prise en compte de ces ressources computationnelles.

À ces deux premières propriétés – la nature particulière des données et la puissance des ressources computationnelles – s'ajoute la recherche de *corrélations*, qui se situe au cœur de l'approche de type *big methods*<sup>53</sup>. Cette approche méthodologique consiste, pour reprendre la conceptualisation de la notion de « modèle » d'après Franck Varenne (2010, 2016), à produire des modèles d'analyse de données en traitant les traces des *big data* avec des méthodes computationnelles<sup>54</sup>. Or, par définition, les modèles d'analyse de données visent à « repérer les corrélations, à résumer et à cribler l'information empirique » (Varenne, 2010, p. 24). Avec l'approche de type *big methods*, l'analyse du degré de dépendance mutuelle entre différents types de données – les traces des *big data* – vient se substituer à l'explication par la mise en évidence de relations de corrélations. Il s'agit, au fond, de repérer des tendances générales sans toutefois s'intéresser à leurs causes. Noortje Marres estime par ailleurs prometteuse cette démarche méthodologique pour entreprendre des analyses de réseaux ou des analyses textuelles (Marres, 2012, p. 149-150).

---

53. Bien que les approches statistiques en sciences sociales procèdent également par la recherche de corrélations, elles se distinguent des *big methods* en ce qu'elles passent par la formulation préalable d'hypothèses qui seront, dans un second temps, validées (ou non) par des calculs. À l'inverse, l'approche de type *big methods* consiste à extraire des hypothèses d'une recherche automatisée de corrélations, effectuée par le biais d'un ensemble d'algorithmes. Il s'agit, de ce fait, d'une démarche radicalement inductive (Boullier, 2015b, 819-822). Voir « *big methods* » dans le glossaire fourni à la fin du présent rapport.

54. La conceptualisation de la notion de « modèle » fait l'objet d'une présentation dans le chapitre II du présent rapport.

Ces techniques d'enquête, dont l'existence précède de loin l'émergence des *big data*, semblent en effet bénéficier pleinement du contexte sociotechnique actuel, dans la mesure où celui-ci autorise la réalisation de recherches fondées sur des analyses de réseaux ou des analyses textuelles à (très) grande échelle.

Pour terminer cette caractérisation de l'approche *big methods* en tant que type idéal, notons que cette stratégie passe par l'appropriation de matériaux exogènes, au sens où les chercheurs en sciences sociales ne sont pas à l'origine de la production des traces des *big data*. Cette approche consiste à emprunter des « données d'occasion » (pour reprendre l'expression ironique de Venturini) à des plateformes en ligne qui sont conçues non pour faciliter le développement de connaissances scientifiques, mais pour répondre à des exigences de valorisation marchande, pour contribuer à des opérations d'optimisation technique ou pour participer à des systèmes de surveillance (Venturini, 2012). Les traces des *big data* sont, autrement dit, fabriquées dans des contextes sociotechniques dans lesquels les impératifs pratiques ont préséance. Ce point n'est pas sans importance, dans la mesure où il soulève la difficile question de la traductibilité de ces traces en données conformes aux normes de la recherche en sciences sociales.

## 1.4 Les « Digital Methods » de la Digital Methods Initiative

### *Quelques remarques préalables au sujet des méthodes numériques*

Les syntagmes « méthodes digitales » et « méthodes numériques » traduisent en français – sans être néanmoins dépourvus d'ambiguïtés – l'expression « digital methods<sup>55</sup> ». Il va sans dire qu'une telle expression, en raison de son niveau de généralité, est susceptible de désigner dans la littérature scientifique des réalités aussi multiples que variables. Ceci étant dit, malgré leur diversité, les stratégies méthodologiques pouvant être regroupées sous l'étendard des *digital methods* semblent partager plusieurs dénominateurs communs : elles servent la réalisation de recherches en sciences sociales ; elles souhaitent se saisir des possibilités qu'autorisent la collecte et l'analyse d'objets nativement numériques ; elles

---

55. Voir dans le glossaire fourni à la fin du présent rapport : « digital vs numérique ».

s'efforcent d'articuler des techniques d'enquête quantitatives et qualitatives. Cette triple caractérisation, dessinant les contours d'un nouvel idéal-type, permet de distinguer les *digital methods* à la fois de l'approche dite *big methods* (cette dernière relevant uniquement d'une démarche quantitative), des ethnographies virtuelles (reposant sur l'adaptation des méthodes ethnographiques au domaine numérique) et, à l'évidence, des méthodes conventionnelles (celles-ci exigeant des chercheurs en sciences sociales qu'ils produisent ou colligent par eux-mêmes leurs données). Dans les deux sections suivantes, nous nous arrêterons plus particulièrement sur les méthodes numériques telles qu'elles sont définies par la *Digital Methods Initiative* (DMI), d'une part, et sur les méthodes numériques « quali-quantitatives », d'autre part<sup>56</sup>.

### **Objet numérisé et objet nativement numérique**

La *Digital Methods Initiative* est une organisation de recherche sous la direction d'un sociologue des Pays-Bas, Richard Rogers, et qui existe depuis 2007. Ce réseau, dont plusieurs collaborateurs travaillent à l'Université d'Amsterdam, aspire à fournir « une contribution à la réalisation de recherches sur ce qui est nativement numérique<sup>57</sup> ». Bien qu'il existe d'autres recherches et publications ayant participé à la définition et au développement de cette approche méthodologique, l'ouvrage *Digital Methods*, signé par Richard Rogers en 2013, reste encore aujourd'hui un

---

56. Nous remercions Antonio Casilli d'avoir attiré notre attention sur une autre classe de méthodes numériques, à savoir l'approche ethnocomputationnelle. Celle-ci a la particularité d'articuler un travail ethnographique à des simulations multiagents (*agent-based modeling*). Antonio Casilli et Paola Tubaro caractérisent cette stratégie méthodologique de la manière suivante : « this methodology [...] is based on three tenants: 1) collecting a small set of qualitative data touching on the target social process; 2) using these data to define a set of behavioural and relational rules for a population of artificial agents implemented in a computer-based environment; 3) generate scenarios that enable exploring the implications of these rules by simulating iteratively interactions among these agents » (Casilli et Tubaro, 2012, p. 7). Nous recommandons la lecture de l'ouvrage de Joshua Epstein et de Robert Axtell pour avoir plus d'information sur les simulations multiagents (Epstein et Axtell, 1996).

57. Il s'agit de notre traduction. *Frequently Asked Questions* (5 avril 2017). Dans *Digital Methods Initiative*. Repéré le 19 mars 2018 à <https://wiki.digitalmethods.net/Dmi/WebHome#FAQ>

incontournable pour en apprécier les spécificités<sup>58</sup>. C'est en effet dans cet ouvrage que se trouve la distinction fondamentale, pour les *digital methods*, entre les objets numérisés et les objets nativement numériques (Rogers, 2013, p. 19-38)<sup>59</sup>. Dans le cadre de cette dichotomie, le terme « objet » ne désigne pas seulement des données. Les *digital methods* s'intéressent, certes, à ces dernières, mais aussi aux méthodes par lesquelles ces données sont fabriquées et manipulées. Partant de là, les objets numérisés renvoient aux données et aux méthodes dont l'existence précède la numérisation de ces objets. À l'inverse, les objets nativement numériques s'identifient aux données et aux méthodes dont l'existence est intrinsèquement liée au numérique. Les objets numérisés se confondent, par exemple, à des photographies ou à des publications scannées, à des informations sociodémographiques collectées à l'aide de formulaires imprimés et saisies par la suite dans des bases de données, à des entretiens semi-directifs menés au moyen de logiciels de voix sur IP, à la mise en place de sondages au format HTML, etc. Les like, les métadonnées, les algorithmes de Google Search, le moteur de recherche sémantique de Facebook (Facebook Graph Search) illustrent quant à eux ce que peuvent être des objets nativement numériques.

Outre qu'elle permet d'entrevoir le champ d'application des *digital methods*, cette distinction s'avère éclairante pour saisir l'originalité de cette stratégie méthodologique en comparaison avec d'autres méthodes pouvant apparaître, à première vue, similaires. Sur la base de cette dichotomie, Richard Rogers définit d'abord explicitement son approche dans un rapport d'opposition avec les *virtual methods*, celles-ci se contentant « d'importer les méthodes standards des sciences sociales et des humanités dans le médium [le numérique] » (Rogers, 2013, p. 19). Il reconnaît en revanche des affinités entre les *digital*

---

58. Pour avoir un aperçu de la diversité des recherches réalisées sous l'égide de la *Digital Methods Initiative*, nous recommandons la consultation du wiki de cette organisation. *Papers and Publications* (14 mars 2018). Dans *Digital Methods Initiative*. Repéré le 19 mars 2018 à <https://wiki.digitalmethods.net/Dmi/PapersPublications>

59. En anglais, Richard Rogers utilise les termes « digitized » et « natively digital » pour rendre compte de cette opposition (Rogers, 2013, p. 19).



*methods*, les *cultural analytics*<sup>60</sup> et les *culturomics*<sup>61</sup>. Pour Richard Rogers, les *cultural analytics* (Manovich, 2007) et les *culturomics* (Michel *et al.*, 2011) appartiennent au domaine, encore émergent, des *digital humanities*. Ces approches entendent renouveler l'intelligibilité de phénomènes culturels en exploitant des méthodes nativement numériques dans le contexte de recherches en sciences humaines et sociales. Cette appropriation de méthodes nativement numériques, matérialisées dans des plateformes et dispositifs en ligne, s'avère être un dénominateur commun aux *digital methods*, aux *cultural analytics* et aux *culturomics*. Les *digital methods* n'ont cependant pas pour finalité d'analyser des données numérisées, ce que font précisément les *cultural analytics* et les *culturomics*. La perspective méthodologique définie par Richard Rogers consiste ultimement à mobiliser des méthodes nativement numériques pour rendre compte de données nativement numériques et non de données numérisées (Rogers, 2015, p. 4-7)<sup>62</sup>.

---

60. Lev Manovich, à l'origine des *cultural analytics*, définit cette approche de la manière suivante : « I define Cultural Analytics as the analysis of massive cultural data sets and flows using computational and visualization techniques. » (2015)

61. Voici la définition des *culturomics* telle qu'elle se trouve sur le site *Culturomics.org* : « Various fields with the suffix «-omics» [...] have emerged in recent years. These approaches tend to rely on (i) hundreds or thousands of people in massive, multi-institutional and multi-national consortia, (ii) novel technologies enabling the assembly of vast datasets containing a very specific type of data, and (iii) the deployment of sophisticated computational and quantitative methods in order to interpret the resulting data. These fields have created data resources and computational infrastructures that have energized biology. The effort to digitize and analyze the world's books has proceeded along these lines. We hope that it will be a forerunner of similar efforts whose goal will be to digitize and analyze other aspects of recorded history, from newspapers, to manuscripts, to incunabula, to artwork. Such efforts would create resources of extraordinary power for scholars and scientists interested in the study of human culture » (*Culturomics*, n.d).

62. Les *culturomics* entreprennent par exemple l'étude de corpus extrêmement volumineux d'ouvrages numérisés avec Google Book (Michel *et al.*, 2011) et les *cultural analytics* l'analyse des propriétés formelles d'un million de pages numérisées de Manga (bandes dessinées japonaises) en exploitant une base de données en ligne (Manovich, 2011).

### ***Le medium web en tant qu'objet, outil et source de connaissance***

Au-delà de cette distinction initiale, les *digital methods* se caractérisent par leur adoption du principe de méthode qui consiste à « suivre le medium » web. Il s'agit en l'occurrence de penser le web en tant que medium. Choisir de parler de « medium » plutôt que de « média » signifie mettre l'accent sur les possibilités cognitives et méthodologiques qu'offre le web en tant que dispositif numérique produisant et conservant les traces relatives aux trajectoires des utilisateurs des plateformes. Il se trouve ici une parenté d'esprit avec les travaux de Marshall McLuhan, penseur pionnier des décennies 1950 et 1960, qui suggérait que le « message, c'est le médium ». McLuhan invitait alors l'analyste à s'attarder aux caractéristiques et aux effets du medium en tant que dispositif agissant dans l'environnement perceptuel, social et cognitif. Ce principe de méthode consistant à « suivre le medium » dessine ainsi les contours d'un programme de recherche original, s'efforçant à la fois de cerner les spécificités des dispositifs en ligne et de se servir de ces derniers pour développer, ultimement, des connaissances sur les changements sociaux et culturels de notre présent.

Dans cette perspective, « suivre le medium » web suppose d'abord d'étudier les méthodes mises à disposition par les plateformes numériques, c'est-à-dire les procédés par lesquels les dispositifs en ligne produisent et gèrent les traces des usagers. Il s'agit, de ce fait, d'envisager le medium en tant qu'objet de recherche. Ce principe implique également d'examiner les possibilités qu'offrent les méthodes natives des plateformes en tant qu'outils méthodologiques dans le cadre de recherches en sciences sociales. Cela signifie, autrement dit, de considérer simultanément les plateformes comme des moyens et des fins. Enfin, « suivre le medium » signifie travailler à la reconfiguration des données et des méthodes nativement numériques pour se les approprier dans le contexte de recherches en sciences sociales<sup>63</sup>. Les objets nativement numériques ne doivent pas, par

---

63. « Follow the methods of the medium as they evolve, learn from how the dominant devices treat natively digital objects, and think along with those object treatments and devices so as to recombine or build on top of them » (Rogers, 2013, p. 5)

le fait même, être intégrés tels quels. En effet, il est essentiel d'employer une démarche réflexive pour les réutiliser et pour construire sur eux des connaissances sur des phénomènes sociaux et culturels en adéquation avec les normes des sciences sociales.

Bien que l'étude des dispositifs en ligne soit une finalité pour les *digital methods*, il ne s'agit en fait que d'une fin intermédiaire et non d'une fin dernière au regard de cette perspective méthodologique. Cette approche s'intéresse, certes, aux spécificités des plateformes web en tant qu'objet de recherche (en analysant, par exemple, des algorithmes, des interfaces, des normes socio-techniques, des fonctions, etc.), mais dans l'optique de formuler *in fine* des diagnostics sur les tendances sociales et culturelles de notre époque. Par ailleurs, cette démarche entend s'affranchir de la distinction entre le réel et le virtuel. D'après le réseau *Digital Methods Initiative*, Internet serait une source légitime pour développer des savoirs sur les changements sociaux et culturels de notre présent se situant au-delà de leurs manifestations concrètes en ligne (Rogers, 2013, p. 21-23). Selon les tenants de cette approche, il apparaît possible de tirer des conclusions, à partir de l'étude d'internet, sur des phénomènes culturels et sociaux traversant les frontières entre le « en ligne » et le « hors-ligne ».

### **Thick Methods**

Professeure au département de *Media Studies* de l'Université d'Amsterdam, Esther Weltevrede conçoit les *digital methods* comme des « *thick methods* », dans la mesure où les *digital methods* semblent réunir des « traditions qualitatives et quantitatives, numériques et interprétatives, universelles et spécifiques » (Weltevrede, 2016, p. 17). Au moins deux arguments viennent soutenir cette affirmation. Le principe de « suivre le medium » requiert d'abord une appropriation des méthodes sous-jacentes aux dispositifs en ligne (les méthodes par lesquelles ceux-ci gèrent, enregistrent ou produisent des traces), en les adaptant aux objectifs de la recherche en sciences sociales. Or, les méthodes des dispositifs en ligne varient considérablement entre elles. Ces méthodes reposent en effet sur des modèles, théories ou représentations très diversifiés, provenant par exemple des sciences sociales, des sciences de la nature ou des sciences informatiques. Les

recherches souscrivant au principe des *digital methods* peuvent donc s'appuyer sur une multiplicité de méthodes, elles-mêmes reposant sur des « conceptions du monde » non moins plurielles. Les *digital methods* ne se réduisent par conséquent ni à une approche qualitative ni à une approche quantitative. Il s'agit, en un sens, d'une perspective composite. Esther Weltevrede établit, de plus, une comparaison entre la démarche analytique des *digital methods* et la « *thick description* », horizon théorique proposé par l'anthropologue Clifford Geertz (1973, 2008). Les *digital methods* incitent, selon elle, les chercheurs à créer des descriptions détaillées et contextualisées de leurs objets d'étude. Cet aspect contribue d'ailleurs à tracer une autre ligne de démarcation entre les *digital methods* et l'approche dite *big methods*. Contrairement aux *digital methods*, l'approche *big methods* n'engage pas la restitution du contexte d'un ensemble limité de traces ou leur observation détaillée<sup>64</sup>.

## 1.5 Les méthodes numériques quali-quantitatives

À l'instar de l'approche préconisée par la *Digital Methods Initiative*, les méthodes numériques dites « quali-quantitatives » entendent contribuer à la recherche en sciences sociales en interrogeant les modalités d'une appropriation raisonnée des données des plateformes en ligne. Les méthodes quali-quantitatives s'apparentent elles aussi à une catégorie particulière de méthodes numériques, conformément à la définition générale que nous en avons donnée plus haut. Elles trouvent leur origine dans le constat d'une configuration sociotechnique inédite, caractérisée par la disponibilité et l'abondance des traces des *big data*, d'une part, et dans l'intention de transformer cette nouvelle donne en opportunité pour les sciences sociales, d'autre part. Trois arguments (au moins) viennent justifier le recours à des méthodes quali-quantitatives : la numérisation généralisée

---

64. Voilà pourquoi certains chercheurs distinguent les données des *big data*, *a priori* sans profondeur et décontextualisées, d'une part, et les « *thick data* » propres aux méthodes travaillant à la « densification » de leurs données, d'autre part (Latzko-Toth, Bonneau et Millette, 2017).

du social, la rigueur dans l'interprétation des phénomènes sociaux et la valorisation de la coproduction des savoirs<sup>65</sup>.

Tommaso Venturini, chercheur au Medialab de Sciences Po, estime aujourd'hui nécessaire la mobilisation des méthodes quali-quantitatives dans les recherches en sciences sociales. Cette appréciation se fonde sur trois observations empiriques : (a) un nombre croissant de phénomènes sociaux se déploient principalement par la médiation des médias numériques ; (b) les pratiques sociales, qu'elles soient en ligne ou hors ligne, s'accompagnent de plus en plus de la production de données numériques ; (c) les dispositifs visant à numériser des données existant sur des supports non numériques se multiplient (Venturini, 2012). Ce diagnostic nous invite, dès lors, à prendre conscience d'une tendance à la numérisation généralisée du social et à en tirer plusieurs conséquences sur le plan méthodologique. Selon Venturini, les méthodes qualitatives ou quantitatives mobilisées jusqu'à présent par les chercheurs en sciences sociales seraient inadéquates pour analyser les phénomènes sociaux actuels. Elles souffriraient de leur incapacité à prendre en charge l'extrême multiplicité et diversité des traces des *big data*, parce qu'elles auraient été développées dans un contexte marqué par la rareté des données sur les phénomènes sociaux (Venturini, Cardon et Cointet, 2014, p. 17). C'est précisément pour cette raison que cet auteur plaide pour la mobilisation d'une approche « capable de réunir les

---

65. Quoique ne traitant pas directement des manières de s'informer en ligne, l'enquête ANAMIA d'Antonio Casilli et Paola Tubaro sur le « web des troubles alimentaires » apporte clairement la preuve de la fécondité des méthodes numériques quali-quantitatives pour la recherche en sciences sociales. Pour avoir plus d'informations à ce sujet : <http://www.anamia.fr/>

avantages des deux types de méthodes», en l'occurrence une méthodologie quali-quantitative (Venturini, 2012)<sup>66</sup>.

Dans un article publié en 2013, Irène Bastard, Dominique Cardon, Guilhem Fouetillou, Christophe Prieur et Stéphane Raux apportent d'autres arguments en faveur des méthodes quali-quantitatives. Ils examinent, dans cette publication, les modalités concrètes de l'articulation entre des techniques d'enquête conventionnelles – quantitatives ou qualitatives – et des techniques computationnelles (Bastard *et al.*, 2013). Les auteurs déconstruisent, en l'occurrence, l'illusion de l'autosuffisance des traces des *big data* dans le contexte de recherches en sciences sociales. Ils montrent que les données numériques, extraites et traitées au moyen de techniques computationnelles ne sont pas significatives par elles-mêmes et requièrent toujours un effort d'interprétation de la part des chercheurs. Or, ce travail implique de recourir à des données provenant de techniques d'enquête conventionnelles. C'est seulement au « contact direct des individus qu'il est possible de conduire des contrôles interprétatifs nécessaires [à l'analyse des données numériques] » (Bastard *et al.*, 2013). La rigueur de l'interprétation des chercheurs en sciences sociales s'avère donc, dans cette perspective, conditionnée par la mise en place de questionnaires, d'entretiens, de groupes de discussion, etc.

Enfin, les techniques d'enquête qualitatives restent incontournables pour rendre possible la participation des acteurs à l'élaboration de connaissances scientifiques concernant leurs propres pratiques sociales. À partir du moment où les chercheurs en sciences sociales accordent une légitimité à la parole des acteurs (Becker, 2004), contre la thèse bachelardienne de la rupture

---

66. Tommaso Venturini défend en réalité une conception des méthodes quali-quantitatives beaucoup plus exigeante et ambitieuse que ce que nous présentons dans cette section. Pour cet auteur, les méthodes quali-quantitatives doivent en effet être une « nouvelle génération de méthodes capable de combler la discontinuité entre le micro et le macro », « de dépasser la distinction micro/macro et de retracer comment les phénomènes globaux sont construits par la coordination d'une multiplication d'actions locales » (Venturini, 2012). Les méthodes quali-quantitatives dont il est question dans notre argumentation s'efforcent, plus modestement, d'articuler des techniques d'enquête qualitatives et quantitatives pour parvenir à exploiter les traces des *big data* dans le cadre de recherches en sciences sociales. Elles n'ont donc pas l'ambition de résoudre méthodologiquement l'un des problèmes majeurs de la sociologie.

épistémologique (Bachelard, 1995), il va de soi que la contribution des acteurs à la construction des connaissances scientifiques ne peut qu’être souhaitable. Le processus de coproduction des savoirs encourage, dans cette logique, l’enrichissement et la complexification des connaissances scientifiques, tout en offrant un levier supplémentaire pour en évaluer la plausibilité.

## 2. Enjeux organisationnels, économiques, politiques et éthiques

En ayant pris connaissance des caractéristiques propres à chacun de ces idéaux-types, nous devrions à présent avoir une vision plus claire des stratégies méthodologiques sur lesquelles peuvent s’adosser les recherches en sciences sociales sur les pratiques informationnelles. Au-delà de leurs propriétés formelles, la mise en œuvre de ces différentes méthodes s’avère traversée par des enjeux relevant autant de l’organisationnel, de l’économique, du politique que de l’éthique. Cette section se propose, en l’occurrence, d’aborder les enjeux les plus fréquemment mentionnés dans la littérature scientifique et dans les discours de nos enquêtés.

### 2.1 Enjeux organisationnels : la collaboration entre chercheurs en sciences sociales et informaticiens

De ce qui précède, on constate une importance grandissante des compétences informatiques dans les méthodes de recherche en sciences sociales. Bien que la mise en place de méthodes conventionnelles ou d’ethnographies en ligne puisse faire l’économie de compétences en informatique, les méthodes computationnelles exploitant les *big data*, les *digital methods* de la *Digital Methods Initiative* et les méthodes numériques quali-quantitatives requièrent pour leur part une solide expertise dans ce domaine (connaissances en algorithmique, en programmation, en modélisation, etc.). Conformément à l’hypothèse de Noortje Marres sur la « redistribution des méthodes » (2012), le projet de collecter et d’analyser des objets nativement numériques encourage donc une ouverture du processus de recherche en sciences sociales à d’autres acteurs. Si, en droit, les chercheurs en sciences sociales peuvent développer des connaissances en informatique pour

assumer tous les aspects d'une méthodologie mobilisant les traces des *big data*, cette éventualité demeure, dans la réalité, peu probable. Le temps nécessaire à l'acquisition de connaissances dans ce domaine risque en effet de décourager même les plus enthousiastes<sup>67</sup>. L'approche de type *big methods*, les *digital methods* de Richard Rogers et les méthodes numériques quali-quantitatives invitent plus vraisemblablement à une collaboration avec des informaticiens (ingénieurs autant que chercheurs) et, de ce fait, à reconsidérer la division du travail scientifique.

Avant d'examiner les modalités possibles de cette coopération, il nous faut « enfoncer une porte ouverte » en contestant l'idée selon laquelle les algorithmes, les programmes informatiques et les applications seraient réductibles à de simples outils, sans réelle incidence cognitive, dans le contexte d'un projet de recherche en sciences sociales. Plusieurs personnes de notre panel ont d'ailleurs fermement critiqué cette représentation, dans la mesure où celle-ci masque les biais intellectuels (et normatifs) que ces dispositifs sociotechniques incorporent<sup>68</sup>. Pour reprendre l'expression de Gaston Bachelard, elle nous empêche de les comprendre comme des « théories matérialisées » (1995, p. 16). Les programmes informatiques, articulant des algorithmes dans différents langages, sont les expressions matérielles de décisions analytiques, à savoir de concepts, de catégories, de règles logiques, etc. Il importe par conséquent de prendre conscience de ces choix analytiques avant d'entreprendre des stratégies méthodologiques visant à exploiter des plateformes en ligne ou des données numériques. Les chercheurs en sciences sociales ne peuvent que difficilement ignorer l'existence de ces biais intellectuels, au risque de compromettre leur capacité à porter un

---

67. La procédure inverse paraît également envisageable (l'acquisition d'une expertise en sciences sociales par un chercheur en informatique, de telle sorte que ce dernier puisse assurer de façon autonome la réalisation d'un projet de recherche en sciences sociales), mais au prix de difficultés similaires.

68. Depuis les premières études sociologiques dites « constructivistes » sur les technologies, publiées par Trevor J. Pinch et Wiebe E. Bijker (1984), nous savons que les dispositifs techniques, loin d'être des instruments neutres, matérialisent des représentations, des aspirations ou des intérêts portés par divers groupes sociaux. Bien que cette approche ne soit pas exempte de critiques, elle présente l'avantage d'exposer avec clarté l'idée selon laquelle la conception des dispositifs techniques ne relève pas uniquement de logiques instrumentales (de la maximisation de leur efficacité).



regard critique, réflexif et rigoureux sur leur propre démarche. Nonobstant les modalités de la collaboration entre chercheurs en sciences sociales et informaticiens, il apparaît impératif que les premiers se dotent de compétences minimales pour discuter avec les seconds et inversement. C'est en ce sens qu'il faut saisir l'appel de certains de nos enquêtés à former les chercheurs en sciences sociales aux fondamentaux de l'informatique, non pas dans l'optique de devenir eux-mêmes des informaticiens, mais dans l'idée de pouvoir prendre des décisions éclairées quant à leurs outils méthodologiques et, ultimement, de s'assurer de l'adéquation de ces dispositifs à leurs problématiques.

La coopération entre chercheurs en sciences sociales et informaticiens peut prendre plusieurs formes. Pour saisir pleinement les différentes modalités de cette association, précisons d'abord ce que signifie l'expression « informaticien » dans le contexte de notre discussion. Ce terme désigne, en réalité, au moins deux statuts professionnels dont la ligne de partage ne fait d'ailleurs pas l'unanimité parmi les informaticiens eux-mêmes : celui de chercheur en informatique et celui d'ingénieur en informatique. Cette distinction professionnelle entre deux métiers mérite d'être soulignée, dans la mesure où elle affecte de multiples façons les conditions de la collaboration entre les chercheurs en sciences sociales et les experts en informatique.

Cette coopération s'identifie parfois à une relation contractuelle entre une maîtrise d'ouvrage (un commanditaire) et une maîtrise d'œuvre (un fournisseur). Contre rémunération, une entreprise s'efforce de répondre aux besoins d'une équipe de recherche en mettant à son service des ingénieurs en informatique. Les chercheurs en sciences sociales définissent alors leurs attentes conformément à une problématique qu'ils déterminent *a priori*. Il va sans dire que les interactions entre ces acteurs demeurent, dans ce modèle, minimales. À rebours de cette approche, la constitution d'un réseau ou d'un groupe de recherche associant des expertises en informatique et en sciences sociales paraît être une option stimulante. Dans ce cadre, la collaboration entre les chercheurs en sciences sociales et les informaticiens – dans les deux sens du terme – semble devoir gagner en intensité et en équilibre. Une telle organisation créerait ainsi les conditions de possibilité d'un véritable échange, d'un partage de compétences et, plus encore, d'un investissement conjoint à toutes

les étapes du projet de recherche, c'est-à-dire de la définition d'une problématique jusqu'à la publication des résultats de l'étude. Cette forme de collaboration encouragerait par là même une parité de participation. Quoique prometteuse intellectuellement, cette organisation présente toutefois des difficultés qu'il serait imprudent de négliger. L'hétérogénéité des cultures scientifiques (et techniques) en sciences sociales et en informatique ne manque pas en effet de menacer la félicité de leur collaboration. Selon nos enquêtés, l'échec est même l'issue la plus probable – mais non inéluctable – de ce type de tentative<sup>69</sup>. Une dernière possibilité, plus rare, passe par l'enrôlement de chercheurs au profil hybride, ayant à la fois des compétences en informatique et en sciences sociales. Ce profil reste cependant exceptionnel, en raison du temps nécessaire à l'acquisition d'une double formation dans des domaines aussi distincts. La relation contractuelle et la constitution d'un réseau (ou d'une équipe) restent par conséquent les options les plus fréquentes.

## 2.2 Enjeux « politiques » de l'accès aux traces des big data

Le recours à l'approche dite *big methods*, aux *digital methods* de la *Digital Methods Initiative* ou aux méthodes numériques quali-quantitatives exige de considérer attentivement la question de l'accès aux traces des *big data*. Contrairement à ce que laisse penser une représentation de sens commun, celles-ci ne sont pas publiques et ne peuvent pas être extraites en toute liberté. En tant que propriétaires des traces des *big data*, les entreprises détentrices des plateformes en ligne, à l'instar de Google, Facebook ou Twitter, jouissent de la prérogative de contrôler l'accès à leurs données numériques. De ce fait, il existe un

---

69. En disant cela, nous ne souhaitons pas laisser conclure en l'impossibilité d'une étroite collaboration entre informaticiens et chercheurs en sciences sociales. L'ethnographe Heather Ford a, par exemple, mené plusieurs projets de recherche – avec succès – sur Wikipédia au côté de deux chercheurs en informatique (2014). Ces expériences sont d'ailleurs à l'origine de fécondes réflexions sur ce qui, au-delà de leurs différences, apparaît commun aux chercheurs en informatique et aux ethnographes (l'adoption de raisonnements inductifs, la recherche de la systématité, etc.). Cette auteure recommande également de s'appuyer sur la complémentarité des compétences des uns et des autres, tout en évitant une division du travail trop rigide, de façon à favoriser une collaboration fructueuse (p. 2-3).

rapport de pouvoir entre les organisations désireuses de mener des recherches en lien avec des données numériques et des plateformes en ligne, d'une part, et les compagnies propriétaires de ces dernières, d'autre part. Voilà pourquoi, en un certain sens, l'accès aux traces des *big data* se révèle être un enjeu politique (Brown, 2002, p. 569-570).

Pour Danah Boyd et Kate Crawford (2012, p. 673-675), le contrôle qu'exercent les entreprises sur l'accès à leurs données peut se traduire de trois façons différentes. Lorsque des chercheurs en font la demande, les propriétaires des plateformes en ligne disposent d'abord du droit d'interdire, à leur discrétion, l'accès à leurs données numériques. Dans ce cas, la réalisation du projet scientifique se voit directement compromise, à moins que son dispositif méthodologique ne soit substantiellement redéfini. Les entreprises ont également la possibilité de vendre des corpus de données numériques. Le prix d'un tel service peut néanmoins s'avérer prohibitif et, là encore, décourager la conduite d'un projet scientifique. Les traces peuvent, enfin, être accordées sans compensation financière aux chercheurs. La tendance actuelle étant plutôt à la valorisation marchande des données numériques, cette option semble toutefois devoir se raréfier à l'avenir (Bastard *et al.*, 2013). Pour accéder gratuitement aux traces, l'interface de programmation des plateformes en ligne – *Application Programming Interface* – est fréquemment privilégiée. Cela étant dit, les entreprises ne mettent pas à disposition la totalité des traces qu'elles colligent. Elles déterminent, à leur convenance, quelles seront les données numériques accessibles, mais aussi pour combien de temps celles-ci le resteront<sup>70</sup>. Une quatrième et dernière option pourrait compléter les analyses de Danah Boyd et Kate Crawford. Les propriétaires des plateformes en ligne ont la possibilité, dans le cadre d'un accord restreint, de mettre gratuitement à disposition (ou à faible coût) un corpus de données numériques auprès d'un groupe de recherche. Une telle décision sera vraisemblablement motivée par des intérêts commerciaux ou financiers. Par rapport à la

---

70. La durée de la disponibilité des données numériques demeure une vive préoccupation pour les équipes de recherche. Celles-ci peuvent en effet se voir interdire, sans avertissement préalable, l'accès à un certain type de traces alors même que leurs travaux sont encore en cours.

troisième option, les traces demeurent inaccessibles à l'ensemble de la communauté scientifique et restent, de ce fait, confidentielles.

Les implications de cette relation de pouvoir ne sont pas sans conséquence d'un point de vue épistémique. Dans la mesure où la reproductibilité demeure, pour bien des chercheurs en sciences sociales, l'un des critères permettant d'évaluer la scientificité des connaissances (Popper, 1973), les problèmes d'accès aux données numériques constituent un obstacle au processus de validation du travail scientifique par les pairs. Même la consultation la plus élémentaire des données collectées et analysées par une équipe de chercheurs se trouve suspendue aux décisions des propriétaires des plateformes en ligne et aux conditions que ceux-ci fixent à l'obtention de leurs traces. Il va sans dire que cet obstacle suscite également des préoccupations quant à la possibilité d'apprécier la plausibilité des études menées par les services de recherche et de développement des entreprises de la Silicon Valley. Le contrôle de l'accès aux données numériques engendre aussi de légitimes interrogations sur les intérêts que les propriétaires de plateformes en ligne poursuivent lorsqu'elles conviennent de transmettre – dans le cadre d'une transaction marchande ou non – des matériaux à des équipes de recherche externes. Ayant pour raison d'être la manipulation des traces des *big data*, l'approche dite des *big methods*, les *digital methods* de la *Digital Method Initiative* et les méthodes numériques quali-quantitatives renforcent inévitablement l'emprise que peuvent avoir les intérêts privés sur les recherches en sciences sociales. Cette hétéronomie (relative) pose par conséquent la question des intérêts sociaux que servent, explicitement ou implicitement, les recherches en sciences sociales se fondant sur de telles stratégies méthodologiques. De ce point de vue, l'exigence de réflexivité que Max Horkheimer assigne à la théorie critique, par opposition à la « théorie traditionnelle » (les sciences positives), ne manque pas de pertinence dans le contexte des recherches en sciences

sociales sur les pratiques informationnelles actuelles (Horkheimer, 1974; 2009)<sup>71</sup>.

Quoiqu'asymétrique, ce rapport de pouvoir entre les propriétaires des plateformes en ligne et les équipes de recherche n'est pas complètement verrouillé. Les chercheurs disposent, encore aujourd'hui, d'un moyen pour contourner le contrôle que les entreprises exercent sur leurs données numériques. Il s'agit de la technique du *scraping*. Noortje Marres et Esther Weltevrede en donnent la définition suivante : « [s]crapers [...] are bits of software code that makes it possible to automatically download data from the Web, and to capture some of the large quantities of data about social life that are available on online platforms like Google, Twitter and Wikipedia » (2013, p. 313). Cette technique, qui semble pouvoir être assimilée à une forme de résistance dans une perspective foucauldienne (2001, p. 1055), est endogène à l'univers de la recherche. Il s'agit initialement d'un procédé conçu pour le domaine du commerce électronique, exigeant par là même un effort d'appropriation par la communauté scientifique (Marres et Weltevrede, 2013, p. 320). En raison de la popularité de ce procédé dans le secteur marchand, il existe une multiplicité d'outils pour effectuer du *scraping*, incluant des addiciels à installer dans des navigateurs, des programmes ou encore des scripts divers. Il va sans dire qu'une telle technique jouit d'une certaine attractivité auprès des chercheurs en sciences sociales, dans la mesure où elle promet une extraction de données numériques sans compensation financière et sans demande d'autorisation auprès des propriétaires des plateformes en ligne. Bien que ces avantages soient réels, la pratique du *scraping* n'est cependant pas dépourvue de problèmes. Les données numériques obtenues par ce procédé requièrent, en premier lieu, un travail conséquent de nettoyage avant d'être exploitables dans le cadre d'une recherche. Elles

---

71. Pour Max Horkheimer, les sciences positives demeurent aveugles aux intérêts sociaux qui les commandent, précisément en raison de leur non-réflexivité. C'est ce qui apparaît clairement dans la célèbre citation suivante : « [L]a science elle-même ne sait pas pourquoi elle met en ordre les faits justement dans telle direction, ni pourquoi elle se concentre sur certains objets et non d'autres. Ce qui manque à la science, c'est la réflexion sur soi, la connaissance des mobiles sociaux qui la poussent dans une certaine direction, par exemple à s'occuper de la Lune, et non du bien-être des hommes » (Horkheimer, 2009, p. 329).

doivent au préalable être interprétées, triées, ordonnées, corrigées et uniformisées. Par ailleurs, cette technique se situe encore aujourd'hui dans une « zone grise » juridiquement parlant. Les dispositions légales entourant la pratique du scraping connaissent des variations importantes en fonction des États et n'encadrent pas toujours clairement l'utilisation de cette technique.

### 2.3 Enjeux économiques : la dépendance des chercheurs au secteur marchand

Le contrôle de l'accès aux données numériques renforce, de plus, les effets des inégalités économiques entre les institutions de recherche. Les organisations les moins dotées financièrement risquent inévitablement d'être pénalisées par les prix parfois très élevés de la vente de corpus de données numériques. À l'inverse, les organisations bénéficiant de conditions économiques favorables peuvent répondre aux exigences mercantiles des entreprises propriétaires de plateformes en ligne et, de ce fait, conduire sans encombre leurs projets de recherche. Le contrôle de l'accès aux données numériques favorise par conséquent la reproduction d'iniquités dans le champ universitaire, entre les universités les plus importantes et celles appartenant à la « périphérie » (Boyd et Crawford, 2012, p. 674)<sup>72</sup>. Enfin, cette dépendance des chercheurs au secteur marchand ne manque pas de susciter de légitimes interrogations sur les intérêts de ces entreprises privées. Par exemple, lorsque celles-ci acceptent de transmettre des jeux de données à des organisations de recherche – gratuitement ou contre rémunération –, la question se pose de savoir pourquoi elles procèdent ainsi. L'hétéronomie du champ universitaire, dans le contexte d'émergence de recherches reposant sur l'approche *big methods*, sur les *digital*

---

72. Ces inégalités économiques concernent l'accès aux données numériques, mais aussi la capacité à mobiliser des personnes ayant les compétences informatiques pour les manipuler et les analyser. Là encore, les institutions universitaires les plus modestes souffriront inévitablement d'un manque de ressources matérielles par rapport aux autres. Les inégalités économiques produisent par là même une double injustice dans le contexte de l'approche des *big methods*, des *digital methods* de la *Digital Methods Initiative* et des méthodes numériques quali-quantitatives.

*methods* ou sur les méthodes numériques quali-quantitatives, pose donc le problème des intérêts économiques et sociaux que servent, consciemment ou non, de telles études scientifiques.

Dans le cas des recherches universitaires concernant Facebook, il existe une division nette entre, d'une part, les groupes universitaires de recherche qui n'ont aucun lien institutionnel direct avec Facebook et, d'autre part, les équipes de recherche dont au moins l'un des membres appartient à la structure de recherche de Facebook (<https://research.fb.com/>) ou a développé un partenariat avec l'entreprise (*fellowship*, projets de collaboration scientifique, bourses doctorales, subventions d'entreprise, etc.). Dans le premier cas de figure, c'est-à-dire celui d'équipes de recherche complètement indépendantes de Facebook, les conditions sont réunies pour que ces chercheurs puissent développer une distance critique suffisante pour produire des enquêtes objectives qui ne seraient aucunement contraintes par les intérêts de l'entreprise. Malheureusement, ces équipes de recherche indépendantes se voient privées d'un accès aux données « qui comptent », c'est-à-dire celles qui sont monétisées et qui constituent le fond de valorisation économique de l'entreprise. Dans le deuxième cas de figure, les équipes de recherche qui ont un lien institutionnel avec Facebook, les chercheurs ont un accès privilégié à des données décisives, mais les enquêtes réalisées sont en quelque sorte contrôlées de manière économique et éditoriale par Facebook.

## 2.4 Problèmes éthiques des recherches en lien avec le numérique

À l'exception des méthodes conventionnelles, l'ensemble des méthodes abordées dans ce chapitre engage de nouvelles considérations éthiques. Si, dans la décennie 1990 et jusqu'au début des années 2000, les recherches dans le champ des *Internet Studies* ne montraient que peu d'intérêt pour les enjeux éthiques de la collecte et de l'analyse des données numériques, l'évolution des discussions scientifiques témoigne, depuis une dizaine d'années, d'une prise de conscience progressive des problèmes éthiques et déontologiques<sup>73</sup> que peuvent engendrer les études

---

73. Pour une vue d'ensemble, voir : Buchanan, 2011.

en lien avec le numérique (Boyd et Crawford, 2012, p. 671-673). C'est la publication, en 2008, d'un article intitulé « Tastes, Ties, and time: a new social network data set using Facebook.com » (Lewis *et al.*, 2008) qui a déclenché les premières polémiques au sujet des méthodes de recherche exploitant des données numériques. Les chercheurs à l'origine de cette publication ont manipulé 1 700 comptes Facebook appartenant à des élèves de *high school*. Or, malgré une démarche d'anonymisation, d'autres chercheurs ont pu aisément retrouver l'identité des adolescents derrière ces 1 700 comptes. Depuis cet événement, la communauté scientifique réfléchit davantage aux questions relatives à la protection de la vie privée des usagers (*privacy*), d'une part, et au consentement des enquêtés, d'autre part.

La protection de la vie privée des usagers ne semblait pas poser de difficulté jusqu'au tournant du 21<sup>e</sup> siècle, dans la mesure où les utilisateurs apparaissaient auparavant moins enclins à dévoiler leur identité sur le Web. Les sites de réseaux sociaux ont néanmoins changé la donne en incitant les usagers à délaissier leur anonymat et à exposer plus nettement leur « identité officielle » (Bastard *et al.*, 2013). La sophistication des techniques de croisement des données numériques, favorisée par la multiplication des plateformes en ligne, a également contribué à l'érosion de l'anonymat en ligne. Rien d'étonnant alors à ce qu'une étude reposant sur des comptes Facebook puisse engendrer, d'un seul coup, des inquiétudes sur l'éthique de la recherche en sciences sociales en lien avec le numérique. Par ailleurs, les données numériques ont été longtemps identifiées à des informations publiques en raison d'une incompréhension quant à leur nature. Le fait qu'elles soient accessibles sans peine, qui plus est en abondance, a pu laisser penser qu'il s'agissait de ressources documentaires exploitables à merci, au même titre que des publications dans la presse quotidienne. Cette confusion entre l'accessibilité et la publicité a donc, elle aussi, participé à la cécité de la communauté scientifique par rapport aux enjeux de la protection de la vie privée des usagers (Latzko-Toth et Proulx, 2013, p. 47-49). L'approche préconisée par Helen Nissenbaum (2004), invitant à respecter les situations sociales dans le cadre desquelles les utilisateurs interagissent en ligne, offre peut-être une solution au problème de la protection de la vie privée des usagers. Cette



auteure redéfinit la «*privacit *» (*privacy*)   la lumi re de son principe «*d'int grit  contextuelle*» (*contextual integrity*). Conform ment   ce dernier principe, les chercheurs en sciences sociales ne devraient recourir   des donn es num riques que dans la mesure o  cette utilisation ne contrevient pas aux attentes normatives – explicites ou implicites – des contextes sociaux dans lesquelles elles circulent. C'est   cette condition que la vie priv e des usagers serait convenablement prot g e.

En sciences sociales, la r alisation de recherches passe usuellement par l'obtention du consentement des enqu t s. Bien que les  tudes en lien avec le num rique ne fassent pas exception dans ce domaine, la collecte et l'analyse des traces des usagers posent plusieurs probl mes sp cifiques (Latzko-Toth et Proulx, 2013, p. 49-50). Pour les besoins de notre argumentation, nous n'en retiendrons que deux. Il n'existe pas, en premier lieu, de r gle permettant de d terminer de mani re syst matique si l'approbation des enqu t s est n cessaire. Chaque cas suppose un traitement particulier, fond  sur la prise en compte de plusieurs variables: «le degr  de publicit  et de privacit  des donn es bien s r, mais aussi la sensibilit  du th me de la recherche, le degr  d'interaction entre le chercheur et les sujets, et la vuln rabilit  de ces derniers». (Latzko-Toth et Proulx, 2013, p. 49). En second lieu, l'identification des personnes aupr s desquelles les chercheurs doivent demander un consentement s'av re particuli rement complexe. Quand bien m me les futurs enqu t s auraient  t  rep r s – malgr  la volatilit  des collectifs en ligne, l' ph m rit  des traces, le relatif anonymat des usagers, etc. –, la probabilit  d'avoir acc s   l'ensemble de ces personnes para t pour le moins r duite.

\* \* \*

L'argumentation d velopp e dans ce chapitre a d fini plusieurs id aux-types de m thode de recherche en lien avec le num rique. En nous inspirant d'une cat gorisation  labor e par Noortje Marres, nous avons distingu  les m thodes conventionnelles qualitatives et quantitatives, l'ethnographie en ligne, les m thodes computationnelles appliqu es aux traces des *big data* et les m thodes num riques (*digital methods*). Ces derni res peuvent  tre d'ailleurs subdivis es en deux familles, les m thodes

numériques de la *Digital Methods Initiative* et les méthodes numériques quali-quantitatives. Après avoir exposé les propriétés spécifiques à chacun de ces idéaux-types, nous avons également souligné les enjeux organisationnels, politiques, économiques et ethniques traversant la mise en œuvre de ces différentes stratégies méthodologiques. Dans le prochain chapitre, nous entendons explorer comment ces dernières sont concrètement mobilisées dans des recherches sur les pratiques informationnelles contemporaines. Nous nous donnerons ainsi la possibilité « de juger sur pièce » des potentialités et des limites de chacune d'entre elles.

# 4

## Potentialités et limites des méthodes de recherche en lien avec le numérique :

### Étude de cas

Dans le chapitre précédent, nous avons opéré une distinction entre quatre idéaux-types de méthode de recherche en lien avec le numérique : les méthodes conventionnelles ; les ethnographies en ligne ; les méthodes computationnelles exploitant les *big data* ; et les méthodes numériques (*digital methods*, elles-mêmes subdivisées entre méthodes numériques de la *Digital Methods Initiative* et méthodes quali-quantitatives). Ce chapitre entend exposer et analyser, pour chacun d'entre eux, une étude de cas jugée représentative de leurs caractéristiques respectives<sup>74</sup>.

---

74. Le choix des études de cas s'est fait, d'une part, en fonction des suggestions de notre panel de chercheurs et, d'autre part, à partir de notre revue de littérature. Les premiers choix ont été discutés au sein d'un comité scientifique (cinq chercheurs) constitué pour surveiller notre démarche d'enquête. Ces premiers choix ont été ajustés en fonction des remarques du comité. Nous nous sommes efforcés de ne retenir que des études de cas qui concernaient les pratiques informationnelles. Il s'agissait de retenir une seule étude de cas pour chaque type de méthodologie.

Les recherches en sciences sociales présentées dans ce chapitre ont toutes pour point commun de s'intéresser centralement aux pratiques informationnelles liées à l'actualité médiatique<sup>75</sup>. Il ne s'agit donc pas simplement de travaux entretenant un lien avec le numérique, mais d'études portant sur des phénomènes au cœur des préoccupations théoriques et pratiques du Centre d'études sur les médias.

Au-delà d'une volonté de donner de la « chair » à notre propos, cette démarche nous donnera aussi l'occasion d'examiner les modalités concrètes de chacune de ces stratégies méthodologiques, tout en mettant l'accent sur leurs potentialités et leurs limites. Nous adopterons à cet effet une grille d'analyse portant sur cinq dimensions : a) les présupposés épistémologiques des recherches ; b) leurs assises théoriques ; c) leurs techniques de collecte et d'analyse des données, sans oublier les articulations entre ces dernières ; d) les résultats obtenus par ces recherches ; e) les limites qu'elles rencontrent au regard de leur stratégie méthodologique. Après avoir étudié séparément chacune de ces recherches, nous effectuerons, dans le chapitre de conclusion, une comparaison entre leurs approches méthodologiques, de manière à cerner, par effets de contraste, le champ d'application où elles apparaissent les plus fécondes.

Pour les méthodes conventionnelles, nous prendrons une étude du Pew Research Center sur les usages des médias socio-numériques et une recherche menée par Guillaume Le Saulnier (Université de Reims Champagne-Ardène) sur la lecture de la presse en ligne, dont les conclusions sont exposées dans un texte datant de 2013. La première viendra illustrer le recours aux méthodes conventionnelles quantitatives, et la seconde le recours aux méthodes conventionnelles qualitatives. Les recherches menées par Coralie Le Caroff (Université Paris II – dorénavant Maître de conférence à l'Université de Lille) dans le cadre de sa thèse de doctorat seront mobilisées pour exemplifier l'ethnographie en ligne. Ce travail porte sur les usages sociopolitiques

---

75. L'étude quantitative du Pew Research Center concerne l'usage des médias sociaux. Cette enquête concerne peut-être plus indirectement les « manières de s'informer », mais force est de constater que l'usage des médias sociaux constitue aujourd'hui une manière de s'informer de plus en plus répandue.

de l'actualité par la médiation de Facebook. Pour les méthodes computationnelles fondées sur l'exploitation des *big data* – se rapprochant des *big methods*, mais tout en s'en distinguant à la fois du fait d'un certain ancrage dans la sociologie –, nous avons choisi une recherche réalisée en Australie par Axel Bruns, Tim Highfield et Stephen Harrington (tous chercheurs au Centre of Excellence for Creative Industries and Innovation, Queensland University of Technology). Ce travail, dont les résultats ont fait l'objet de plusieurs publications en 2013, concerne le partage d'informations d'actualité par la médiation de Twitter. Pour les méthodes numériques, nous nous focaliserons sur les méthodes numériques quali-quantitatives en nous intéressant à une recherche réalisée par Irène Bastard (alors stagiaire à Orange Labs et doctorante à Télécom ParisTech, aujourd'hui chercheure à la Bibliothèque nationale de France). Cette dernière recherche porte sur les manières de s'informer par Facebook, avec une focalisation sur les usagers adolescents.

## 1. Les méthodes conventionnelles

Dans la mesure où les méthodes conventionnelles existent depuis plusieurs décennies, qu'elles font l'objet d'une pléthore de publications et sont enseignées dans la plupart des formations universitaires en sciences sociales, nous n'aborderons que succinctement les études dans cette première section. Précisons que nous appliquerons quand même systématiquement notre grille d'analyse aux deux cas relevant de cette première section. Nous développerons ensuite les trois études de cas subséquentes.

### 1.1 Une étude quantitative des usages des médias socionumériques

Notre première étude provient du Pew Research Center. Il s'agit d'une enquête, réalisée annuellement auprès d'un public américain, sur les usages des médias sociaux. Les conclusions de cette dernière sont développées dans un rapport intitulé *Social Media Use in 2018* (Smith et Anderson, 2018). Bien que ce document comporte une notice méthodologique, il faut consulter un autre espace du site web du Pew Research Center (2018a) pour connaître en détail les techniques de collecte et d'analyse des données quantitatives sur lesquelles repose leur étude.

*Présupposés épistémologiques.* Au regard de la mission de cette organisation « non partisane » (c'est ainsi que se définit le Pew Research Center), les auteurs de cette enquête n'apportent que peu de précisions concernant les présupposés épistémologiques et théoriques de leur travail. La présentation que donne le site web du centre suggère toutefois une adhésion implicite à une posture que l'on pourrait qualifier de positiviste, dans la mesure où cette organisation entend conduire des recherches empiriques pour donner des bases factuelles aux dialogues et aux décisions publiques<sup>76</sup>. Cette conception de la connaissance scientifique trouve l'une de ses premières formulations systématiques dans l'épistémologie de Rudolph Carnap (1973), épistémologie appartenant au courant du « positivisme logique ». Au risque de caricaturer sa pensée, notons que ce membre du cercle de Vienne défendait dans les années 1930 l'idée selon laquelle la scientificité des connaissances relèverait ultimement d'une exacte correspondance entre des énoncés d'observation et des faits. Cette adéquation entre les uns et les autres constituerait, dans cette perspective, une base empirique à la fois « irrécusable et invariable d'où part et où revient toute connaissance scientifique digne de ce nom » (Soler, 2009, p. 101). Bien que cette affirmation exigerait d'être grandement raffinée, du fait notamment que Carnap réfléchissait avant tout au domaine des sciences naturelles, les études quantitatives menées par le Pew Research Center, relevant du domaine des sciences sociales, ne nous semblent toutefois pas sans lien avec cette posture épistémologique.

Il faut admettre cependant que les « savoirs positifs » produits par le Pew Research Center peuvent ouvrir à une multitude d'interprétations et de questionnements de la part des chercheurs qui exploitent ces données descriptives. La présentation de données brutes tirées des enquêtes par sondage ne signifie pas nécessairement que ces données « parlent d'elles-mêmes » ou qu'elles se suffisent. Ce que l'on peut dire c'est que la présentation des résultats de ces enquêtes quantitatives peut permettre la production d'analyses différenciées selon la problématisation qui est mise de l'avant par l'une ou l'autre équipe de recherche. Il

---

76. Selon les termes du site web du *Pew Research Center*, la mission de cette organisation consiste non pas à prescrire, mais à informer (2018b)

apparaît nécessaire de multiplier de telles études – surtout des analyses statistiques poussées qui ne se contentent pas d’une visée sommairement descriptive – afin de pouvoir déceler des tendances ou avancer des explications avec une solidité scientifique.

*Assises théoriques.* Dans la logique de ce qui précède, les enquêtes annuelles de cette organisation ne procèdent ni de l’adoption *a priori* d’un dispositif théorique ni de la construction d’une problématique au sens où les chercheurs en sciences sociales l’entendent habituellement. Elles reprennent, en réalité, les paramètres démographiques du bureau du recensement américain (U.S. Census Bureau), par exemple l’âge, le niveau d’études, le genre, la race, le lieu de naissance, la région domiciliaire et la densité de population<sup>77</sup>, mais aussi la division des plateformes en ligne telle qu’elle est établie par l’industrie (médias sociaux, plateformes de partage de vidéo, etc.). Ces enquêtes annuelles ne reposent donc pas sur une conceptualisation propre des phénomènes dont elles s’efforcent de rendre compte, mais empruntent plutôt directement leurs catégories d’analyse à l’administration américaine ou au secteur marchand.

*Techniques de collecte et d’analyse des données.* Ces enquêtes, parce qu’elles sont répétées chaque année depuis 2012, permettent d’entrevoir les contours de l’évolution des usages des médias socionumériques aux États-Unis. Sur le plan des techniques de collecte de données, elles s’appuient sur la réalisation d’un sondage téléphonique. Celui-ci consiste à contacter aléatoirement des enquêtés sur leur téléphone cellulaire ou résidentiel. Les personnes répondant à un appel sur leur téléphone cellulaire se voient directement proposer de participer au sondage. Dans le cas d’une réponse à un appel sur un téléphone résidentiel, c’est la personne la plus jeune du foyer qui se trouve sollicitée. Dans le cadre du rapport de 2018, l’échantillon ne comportait pas moins de 2 002 adultes âgés au minimum de 18 ans et répartis sur l’ensemble du territoire américain. Pour être représentatif, cet échantillon a fait l’objet d’une pondération en fonction du portrait de la population états-unienne dressé par le bureau du recensement américain. C’est notamment pendant cette

---

77. Le bureau du recensement américain divise le territoire états-unien en quatre régions, à savoir le Nord-Est, le Sud, l’Ouest et le Midwest.

opération de pondération que les paramètres démographiques de cette administration sont mobilisés par les équipes du Pew Research Center.

Le sondage est lui-même effectué sur la base de 17 questions regroupées en 3 ensembles: le premier concerne les usages des médias sociaux, le deuxième la fréquence d'utilisation de ces derniers et la troisième l'évaluation subjective de la difficulté à se séparer de différentes technologies de communication (la télévision, le téléphone cellulaire, internet, les médias sociaux). Pour analyser les données collectées par l'intermédiaire de ce questionnaire, les auteurs du rapport ont mobilisé des techniques statistiques. Les explications méthodologiques du Pew Research Center font apparaître une volonté de repérer, sur le plan des usages des médias socionumériques, des corrélations statistiquement significatives entre des groupes de répondants (2018a), ces groupes étant déterminés en fonction des critères du bureau du recensement américain.

*Les résultats obtenus.* Plusieurs enseignements émanent de la lecture de ce rapport. Elle nous apprend d'abord que la grande majorité des Américains utilise fréquemment et massivement Facebook et Youtube. Ce rapport révèle également que les adultes les plus jeunes, appartenant à la catégorie des 18-24 ans, font usage d'un large éventail de médias socionumériques. En plus des deux plateformes en ligne mentionnées ci-dessus, cette tranche d'âge utilise de manière fréquente Snapchat, Instagram et, dans une moindre mesure, Twitter. Cette enquête dévoile, en ce sens, un rapport inversement proportionnel entre l'âge et le nombre de médias socionumériques faisant l'objet d'une utilisation fréquente par les répondants. Enfin, elle met en lumière un clivage au sujet de l'appréciation de la difficulté à délaisser les plateformes en ligne. Bien qu'une majorité estime peu difficile de rompre avec les médias socionumériques, une importante minorité affirme le contraire. Les répondants appartenant à la catégorie des 18-24 ans apparaissent particulièrement rétifs à l'éventualité de mettre un terme à leurs usages des plateformes en ligne.

*Les limites du dispositif méthodologique.* Bien qu'il soit envisageable d'entreprendre une critique interne du protocole d'enquête du Pew Research Center, du point de vue des normes



de la recherche quantitative en sciences sociales, nous souhaitons plutôt considérer les limites des méthodes conventionnelles quantitatives pour rendre compte des usages des médias socio-numériques. Ces études, en quantifiant les utilisations des plateformes en ligne, permettent de décrire plusieurs aspects de leur évolution depuis maintenant six ans. Au regard de la mission que se donne cette organisation, elles apportent bien des informations – certes en quantité restreinte – pouvant servir de bases factuelles à d'autres analyses. Ces enquêtes présentent l'avantage de produire rapidement et de manière récurrente des données originales publiques, relativement compréhensibles et « visualisables », mais celles-ci comportent néanmoins le risque d'être mal interprétées ou sur-simplifiées. De telles enquêtes permettent aussi d'observer des pratiques sur plusieurs années ; elles peuvent fournir une base de données susceptible d'être associée à des travaux de comparaisons internationales<sup>78</sup>, y compris avec des approches mixtes. Par ailleurs, ces techniques de collecte et d'analyse de données n'offrent que peu de prise pour comprendre les significations que prêtent les répondants à leurs propres usages ou pour saisir ces pratiques sociales dans leur contexte. Parce qu'elles s'appuient sur une méthode quantitative, ces études souffrent de cécité par rapport aux logiques d'usage, que ces logiques soient culturelles, sociales, psychologiques ou encore politiques. Nous retrouvons de ce fait une préoccupation soulevée par nos enquêtes (mentionnée dans le deuxième chapitre) concernant la nécessité de recourir à des dispositifs méthodologiques permettant à la fois de mesurer, de décrire et de comprendre les usages des médias siconumériques.

## 1.2 Une étude qualitative sur la lecture de la presse en ligne

Notre seconde étude de cas trouve sa source dans un chapitre publié par Guillaume Le Saulnier, en 2013, intitulé « La lecture de la presse en ligne. L'appropriation des contenus d'actualité au défi de la technique ». Le Saulnier examine, en mobilisant une méthode conventionnelle qualitative, la reconfiguration des

---

78. Tel est le cas du *Digital News Report* de la Reuters Institute avec lequel coopère le CEM : <http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475>

manières de lire dans le contexte du développement de la presse en ligne. Il prend pour objet les modalités de consultation des contenus journalistiques en considérant les évolutions des « supports matériels, des fonctionnalités techniques et des conventions de lecture » (Le Saulnier, 2013, p. 58). Cette recherche entend se démarquer de la littérature scientifique en s'intéressant non pas au « journalisme participatif » (à savoir l'une des formes les plus sophistiquées de l'usage de la presse en ligne), mais aux modes de consommation « ordinaires » de l'actualité médiatique. Elle s'en distingue également par son recours à des entretiens avec des lecteurs alors que le questionnaire demeurait une technique d'enquête qui semblait fréquemment exploitée dans la littérature savante à ce moment-là (Le Saulnier, 2013, p. 58).

*Présupposés épistémologiques.* L'auteur se montre peu prolix concernant les présupposés épistémologiques de sa recherche. Cette absence se comprend aisément dans la mesure où cette dimension épistémologique avait été abordée plus globalement dans l'introduction de l'ouvrage dirigé par Josiane Jouët et Rémy Rieffel, et dont est tiré ce chapitre. Ceci étant dit, Guillaume Le Saulnier apporte au moins deux précisions permettant de conjecturer sur les présupposés épistémologiques de son étude. Il prend tout d'abord explicitement position contre ce qu'il identifie comme étant « l'illusion positiviste » (2013, p. 71). Il conçoit en effet les entretiens en tant qu'outil adapté à la collecte et à l'analyse des *déclarations* d'un ensemble d'acteurs au sujet de leurs rapports à la presse en ligne. Ce faisant, il conteste l'idée selon laquelle cette technique d'enquête offrirait la possibilité de connaître la « réalité elle-même » de leurs rapports. Il rappelle, autrement dit, que les entretiens servent à produire *des comptes rendus* d'une pratique sociale et non à restituer la réalité effective de cette dernière. Au-delà de cet indice, on peut dire du cadre théorique de cette recherche qu'il témoigne d'un désir d'interpréter les manières de lire la presse en ligne. En citant Philippe Mallein et Yves Toussaint (1994), Guillaume Le Saulnier dit vouloir s'intéresser aux « significations d'usage » (2013, p. 59), c'est-à-dire au sens que les acteurs attribuent à leurs usages des contenus d'actualité sur internet. Or, l'analyse des significations requiert, même minimalement, de la part du savant une certaine activité interprétative. En conséquence, les assises théoriques de

cette étude suggèrent son inscription dans le domaine des sciences interprétatives<sup>79</sup>.

*Assises théoriques.* Cette recherche se situe théoriquement à la jonction de la sociosémantique et de la sociologie des usages. Bien qu'il mette en question la pertinence, dans le contexte actuel, du développement de la presse en ligne, Guillaume Le Saulnier prend pour point de départ la thèse, énoncée par Umberto Eco (1985), selon laquelle la lecture reposerait sur l'existence de conventions partagées entre les émetteurs et les récepteurs. L'auteur reprend notamment le concept, élaboré par Eliseo Veron, de « contrat de lecture<sup>80</sup> ». Ce repère bibliographique atteste ainsi d'une référence (même lointaine) à la sémiotique pour cette enquête essentiellement sociologique sur les manières de lire la presse en ligne. L'enracinement dans la sociologie des usages est explicite, en plus d'être confirmé par les nombreuses références à des publications dans ce champ, lorsque l'auteur convoque la notion de « significations d'usage » (Mallein et Toussein, 1994). Cette notion met en effet en lumière les « représentations et valeurs [qui] s'investissent dans l'usage d'une technique ou d'un objet » (Chambat, 1994, p. 262). Guillaume Le Saulnier mobilise, pour terminer, plusieurs contributions des sciences de l'information et de la documentation pour conceptualiser ce qu'il entend par « lecture ». Il précise ainsi intégrer trois dimensions pour penser le lire, à savoir « l'expérience sensori-motrice », « le décodage du matériau signifiant » et « l'actualisation des informations et des connaissances » (Le Saulnier, 2013, p. 57).

---

79. Il s'agit ici d'une lecture « philosophique » qui est posée sur ce travail sociologique. Ce constat pourrait paraître en effet superflu aux yeux du sociologue, car pour ce dernier, toutes les recherches en sciences sociales sont par définition « interprétatives ». Nous reviendrons plus longuement sur cette notion de « sciences interprétatives » à l'occasion de notre analyse des travaux de Coralie Le Caroff.

80. Dans un entretien avec Eliseo Veron, pour la revue *Quaderni*, Pierre Gonzales définit le « contrat de lecture » de la manière suivante : [il s'agit] d'un dispositif d'énonciation, qui diffère selon les organes de presse, où est mise en forme la relation de lecture que propose le producteur du discours, en l'occurrence l'éditeur et l'équipe rédactionnelle, à ses lecteurs. Le contenu compte, mais la façon de dire prime pour ce qui est d'établir une relation particulière, si possible durable, un lien affectif ou privilégié avec le lecteur. » (1996, p. 51)

*Techniques de collecte et d'analyse de données.* D'un point de vue méthodologique, cette étude se fonde sur la réalisation de 20 entretiens. Dix-sept entretiens ont été menés avec des lecteurs de la presse en ligne et trois avec des contributeurs actifs, produisant par eux-mêmes des contenus d'actualité sur les plateformes journalistiques. L'auteur n'apporte cependant aucune précision quant à la nature de ces entretiens. Les extraits contenus dans le chapitre témoignent néanmoins d'une certaine spontanéité dans les échanges entre l'enquêteur et l'enquêté (les interventions du premier étant indiqué en italique dans les extraits des entretiens). Nous pouvons, dès lors, en déduire qu'il s'agissait d'entrevues semi-directives ou non directives. L'analyse des données collectées a été effectuée de manière « verticale et transversale » (Le Saulnier, 2013, p. 59). Si l'analyse verticale consiste à prêter attention aux thèmes les plus saillants d'un seul entretien, l'analyse transversale procède par l'examen des thèmes circulant dans un ensemble d'entretiens.

Par ailleurs, les personnes ayant participé à cette recherche ont été recrutées de proche en proche. L'enquêteur a, de ce fait, demandé aux enquêtés de lui suggérer d'autres individus à contacter pour poursuivre l'investigation. Quoiqu'efficace, ce mode de recrutement présente le désavantage de favoriser la constitution d'un corpus homogène au regard des caractéristiques sociodémographiques des acteurs. Guillaume le Saulnier fait mention de ce biais dans son argumentation, en constatant une surreprésentation d'individus appartenant « aux fractions urbaines les mieux dotées en termes de certification scolaire et de niveau de revenu, à commencer par les cadres supérieurs et les professions intellectuelles » (2013, p. 58-59). L'existence de ce biais se voit corroborée en annexe du chapitre, dans un tableau récapitulant les propriétés sociodémographiques de chacun des acteurs<sup>81</sup>.

*Résultats obtenus.* Les contributions de ce chapitre sont triples. Elles apportent d'abord des « descriptions » des pratiques infor-

---

81. Guillaume le Saulnier émet cependant l'hypothèse selon laquelle cette surreprésentation des cadres supérieurs et des professions intellectuelles serait conforme à la réalité sociologique du lectorat de la presse en ligne au moment de sa recherche.

mationnelles des enquêtés (Le Saulnier, 2013, p. 59), en déterminant comment ceux-ci procèdent pour sélectionner leurs supports médiatiques en ligne, en précisant les modalités par lesquelles ils articulent leur lecture des contenus d'actualité sur internet avec d'autres offres médiatiques, mais aussi en situant ces acteurs dans l'espace social (Le Saulnier, 2013, p. 59-63). Dans un second temps, l'auteur montre comment la lecture de la presse connaît des complications, sinon des perturbations, dans le contexte du numérique (p. 63-70). Bien qu'il existe des variations entre les différentes plateformes médiatiques, les manières de lire la presse en ligne s'avèrent globalement moins évidentes qu'avec la presse imprimée au regard de « la sélection, du décodage et de l'approfondissement des contenus » (p. 64). Ces difficultés s'expliquent, en l'occurrence, par une relative inadéquation des conventions de lecture de la presse imprimée avec les contenus d'actualité en ligne. Enfin, dans un troisième temps, Guillaume Le Saulnier rend compte de l'appropriation différenciée des dispositifs d'autopublication, c'est-à-dire des fonctionnalités permettant aux enquêtés de contribuer à la production, à la modification et à la circulation de contenus d'actualités (p. 70-75). Il démontre, ce faisant, une corrélation entre le capital culturel des enquêtés et leur inclination à un usage participatif des plateformes médiatiques. Il révèle également une consolidation du contrat de lecture, « du lien établi entre un support et ses lecteurs » (p. 70) avec ces dispositifs d'autopublication.

*Limites du dispositif méthodologique.* Cette enquête, en s'appuyant sur la réalisation de 20 entretiens, fait face à des limites sur le plan de la généralisation possible de ses résultats et de la représentativité de son corpus. Si cette investigation se montre féconde pour saisir les manières de lire la presse en ligne d'un petit nombre d'enquêtés, le périmètre du champ de validité de ses résultats demeure inconnu. Cela signifie non pas que les conclusions de cette étude ne pourraient être étendues à d'autres acteurs, mais que le « dispositif de preuve » de la recherche ne permet pas de le déterminer<sup>82</sup>. En outre, le corpus de cette

---

82. Nous développerons cette question des possibilités de généralisation des résultats d'enquête, en examinant les limites du dispositif méthodologique de la recherche de Coralie Le Caroff.

enquête semble souffrir d'un problème de représentativité de l'échantillon. Guillaume Le Saulnier a conscience de ce biais dans sa démarche méthodologique et l'explique clairement dans son chapitre. Nous pouvons ainsi nous demander si son investigation parviendrait aux mêmes résultats si les cadres supérieurs et les professions intellectuelles étaient moins surreprésentés.

## **2. Une ethnographie en ligne des usages sociopolitiques de Facebook**

Pour donner une illustration à l'ethnographie en ligne, nous avons sélectionné la thèse de doctorat de Coralie Le Caroff, soutenue en 2015, intitulée « Les usages sociopolitiques de l'actualité en ligne : s'informer, partager et commenter sur Facebook » (2015a). Plusieurs publications émanent plus ou moins directement de cette recherche, les unes mettant l'accent sur des conclusions spécifiques (Le Caroff, 2015b, 2018) et les autres explicitant certains aspects de sa démarche méthodologique (Jouet et Le Caroff, 2016; Béliard et Le Caroff, 2018). L'auteure résume l'objet de sa thèse de doctorat de la manière suivante : « [elle] porte sur les usages sociopolitiques de l'actualité en ligne par les profanes sur Facebook » (Le Caroff, 2015a, p. 7). Cette recherche s'intéresse, pour synthétiser, aux pratiques informationnelles d'un ensemble d'utilisateurs de Facebook ayant la particularité d'être ni des professionnels de la politique ni des militants. Elle examine leurs manières de « s'informer, d'échanger et de commenter les nouvelles » (2015a, p. 7) à la lumière des caractéristiques sociotechniques de ce réseau social, des relations qu'entretiennent ses utilisateurs à la politique, mais aussi des rapports de genre travaillant leurs interactions discursives. Partant de là, ce travail apparaît plus que pertinent pour juger sur pièce du recours à l'ethnographie virtuelle dans les recherches en sciences sociales sur les pratiques informationnelles.

*Présupposés épistémologiques.* D'un point de vue épistémologique, cette recherche se trouve à l'opposé d'une posture dite « positiviste ». Elle ne prétend pas décrire le réel en formulant des énoncés d'observation, dont la validité serait sanctionnée par leur correspondance à des faits. Elle ne souscrit pas non plus

à l'ambition de produire des explications nomologiques<sup>83</sup> pour rendre compte de relations de causalité entre des phénomènes. Elle entend en revanche construire des interprétations pour restituer le sens d'un ensemble de pratiques sociales. Au regard de l'épistémologie, cette thèse de doctorat en sociologie relève ainsi, au sens large, du domaine des sciences interprétatives (au même titre que l'histoire, la psychanalyse, etc.). Les fondements épistémologiques de cette recherche font d'ailleurs l'objet d'une explicitation dans le troisième chapitre de la thèse. Dans celui-ci, Coralie Le Caroff manifeste clairement son attachement à une « approche par les significations d'usage » (2015a, p. 86), laquelle présente l'avantage « d'articuler le versant pragmatique et compréhensif » (p. 86), à savoir la description fine des pratiques des acteurs et des situations dans lesquelles celles-ci ont lieu, d'une part, et l'analyse des significations conférées par ces mêmes acteurs à leurs propres pratiques, d'autre part. Partant de là, elle insiste sur la nécessité pour le chercheur d'interpréter « les discours, les critiques, les justifications “ordinaires” pour comprendre le sens donné aux pratiques des technologies » (p. 86). Or, ce type d'activité cognitive, tournée vers l'interprétation, ne saurait avoir pour critère de validité l'adéquation entre un énoncé d'observation et un fait, autrement dit, entre une représentation et une réalité de référence. C'est plutôt ici la plausibilité qui donne une valeur de scientificité aux interprétations des sciences sociales (Olivier de Sardan, 2008, p. 7-38; Soler, 2009, p. 87-91)<sup>84</sup>. Plus une interprétation apparaît crédible

---

83. C'est-à-dire : les rapportant à des lois, à des régularités, etc.

84. Pour l'anthropologue Jean-Pierre Olivier de Sardan, la plausibilité n'est rien de moins que le propre des connaissances relevant de méthodes qualitatives. Il définit cette approche épistémologique de la manière suivante : « les connaissances ainsi produites ne sont rien d'autre que des approximations plausibles, c'est-à-dire des représentations savantes qui ont pour ambition de rendre approximativement et plausiblement compte des réalités de référence. [...] On a affaire, avec l'enquête de terrain propre à l'anthropologie ou à la sociologie qualitative, à un entrelacs complexe d'ordres de grandeur, de tendances, de descriptions, d'illustrations, de cas significatifs, de discours et de représentations “locales”, d'hypothèses souples, d'interprétations prudentes, de théories locales, de généralisations plus ou moins assurées, le tout pris dans des constantes de variations d'échelle et de perspective. Avec cette mosaïque quelque peu hétéroclite de données commentées et interprétées, nous sommes bien dans de l'à-peu-près. Mais cet à-peu-près n'a rien (ne devrait rien avoir) d'un n'importe quoi. » (Olivier de Sardan, 2008, p. 11-12).

au regard de ce qui fait consensus dans la communauté scientifique, et moins celle-ci exige d'être appuyée sur des éléments de preuve.

*Assises théoriques.* Cinq notions constituent le noyau théorique de la thèse de doctorat de Coralie Le Caroff. Elles s'inscrivent dans la continuité de la posture épistémologique identifiée ci-dessus, dans la mesure où elles invitent à la construction de connaissances plausibles, fondées sur l'interprétation de données empiriques. Par ailleurs, si nous parlons ici de « notions », et non de « concepts », c'est parce qu'elles ne sont pas organiquement liées au sein d'une théorie systématique et unifiée. Elles appartiennent, en réalité, à différents courants de pensée ou à diverses traditions de recherche en sciences sociales. Les notions « d'usage » et de « double médiation » s'avèrent être les premières d'entre elles (Le Caroff, 2015, p. 80-85). L'une et l'autre découlent de la sociologie des usages (Jauréguiberry et Proulx, 2011 ; Proulx, 1994). Contre une approche déterministe de la technique, la notion « d'usage » reconnaît aux usagers une certaine autonomie – contrairement à la figure du récepteur passif dont le comportement serait dicté par les prescriptions des concepteurs – et prend en compte autant la consommation, l'utilisation que l'appropriation des dispositifs techniques par les usagers. La notion de « double médiation » (Jouet, 1993), qui n'est pas l'exclusivité de la sociologie de l'innovation (Akrich, Callon et Latour, 2006 ; Coutant, 2015), permet quant à elle de penser les technologies comme des dispositifs sociotechniques, c'est-à-dire comme des réalités conditionnant les pratiques sociales tout en étant elles-mêmes conditionnées par les pratiques sociales. Coralie Le Caroff convoque également la notion « d'informalité » dans sa thèse de doctorat, laquelle est extraite des recherches sociologiques de Cas Wouters (2007). L'informalité, à rebours des attentes normatives de la première théorie habermassienne de l'espace public (1992), offre la possibilité d'appréhender « les manières de parler, d'exprimer les opinions selon les formats qui renvoient à ceux de la conversation ordinaire, du bavardage, de l'expression des émotions et des affects, de la polémique, qui peuvent toutefois supporter une critique profane » (Le Caroff, 2015, p. 90). L'auteure reprend aussi la notion « d'identité numérique » dans son cadre théorique, laquelle se situe à la confluence de



recherches en sciences sociales sur les modalités de présentation et d'exposition de soi en ligne, sur les cultures expressives (Allard et Vandenberghe, 2003) ou encore sur l'intersubjectivation technique (Voirol, 2011). La notion de « genre » vient compléter les assises théoriques de cette recherche (Béréni, Chauvin, Jaunait et Revillard, 2008) Cette dernière est élaborée de manière à intégrer quatre dimensions, à savoir une distinction entre le genre comme construction sociale et le sexe comme réalité biologique; une conception dialectique du genre, le féminin n'existant que dans sa relation au masculin (et réciproquement); une identification des rapports sociaux de genre à des rapports de domination; une intégration des perspectives intersectionnelles, visant à assumer les interrelations entre les rapports de genre et les rapports de classe (Le Caroff, 2015, p. 98).

*Techniques de collecte et d'analyse des données.* La stratégie méthodologique de cette thèse de doctorat comporte deux volets. Elle conjugue la mise en place d'une observation ethnographique en ligne à la réalisation d'entretiens semi-directifs<sup>85</sup>. La méthode de cette recherche se montre tout à fait pertinente au regard de ses trois hypothèses de départ, dans la mesure où elle autorise la collecte et l'analyse de données permettant d'interroger la nature sociotechnique des usages de l'actualité par Facebook (I), la participation accrue des femmes à des discussions politisées par ce réseau social (II) et, enfin, l'épanouissement de pratiques expressives, contribuant à la construction de soi, par la médiation de cette plateforme en ligne (III). Par ailleurs, l'observation ethnographique en ligne et les entretiens ne constituent en aucun cas deux moments séparés dans l'enquête. Ils demeurent articulés entre eux, formant une stratégie méthodologique cohérente.

L'observation ethnographique en ligne est qualifiée « de méthode artisanale d'immersion » par l'auteure, par opposition

---

85. Dans le chapitre méthodologique de sa thèse de doctorat, Coralie Le Caroff opère une distinction entre l'observation ethnographique en ligne, d'une part, et la réalisation d'entretiens semi-directifs avec les enquêtés, d'autre part. Il s'agit, en réalité, de deux techniques d'enquête concourant à la même stratégie méthodologique, en l'occurrence ce que nous appelons « une ethnographie en ligne ».

aux méthodes computationnelles reposant sur l'exploitation des *big data* (Le Caroff, 2015, p. 103). Contrairement à ces dernières, exigeant des dispositifs informatiques sophistiqués (logiciels d'aspiration de données, etc.) et des ressources financières non négligeables, l'observation ethnographique en ligne offre, selon elle, la possibilité d'étudier de manière innovante les usages des médias socionumériques avec des moyens beaucoup plus modestes<sup>86</sup>. Pour Coralie Le Caroff, cette démarche présente également l'avantage de respecter la richesse et la complexité des usages des plateformes en ligne. Nous retrouvons là, en filigrane, la question épistémologique des limites du pouvoir d'intelligibilité des recherches en sciences sociales fondées exclusivement sur la mesure. Par ailleurs, la dimension immersive de cette méthode se comprend sans difficulté au regard de la durée relativement longue de l'observation (plusieurs mois au minimum). Cette immersion s'avère être une nécessité pour l'auteure, en ceci qu'une compréhension – au sens épistémologique du terme – des logiques sociales et techniques de la participation par Facebook requiert une appropriation authentique de ce média socionumérique. Or, cette appropriation ne peut faire l'économie d'une pratique régulière, éclairée et approfondie du réseau social. Le chercheur doit devenir un « usager averti » (Le Caroff, 2015, p. 109). Dans le cadre de cette recherche, l'immersion n'est par conséquent rien de moins que la condition de possibilité de la compréhension.

L'observation ethnographique en ligne avait pour corpus 2 types d'espaces numériques sur Facebook, en l'occurrence 5 pages appartenant à des organisations médiatiques « conventionnelles », d'une part, et 19 profils personnels d'utilisateurs de Facebook, d'autre part. Les pages des médias sur Facebook ont fait l'objet d'un travail de terrain pendant deux périodes distinctes, chacune ayant été l'occasion d'examiner plusieurs fils

---

86. Il ne faut toutefois pas négliger le fait qu'une démarche de type ethnographique demande un investissement en temps considérable.

de discussion directement liés à l'actualité médiatique<sup>87</sup>. L'auteure a donc étudié pas moins de 1 285 commentaires, dans l'idée de documenter « les espaces participatifs autour de l'actualité sur le web social » (Le Caroff, 2015, p. 105). Les profils personnels des usagers de Facebook ont quant à eux été observés tout au long du processus de la recherche. En procédant de la sorte, des données ont été collectées sur la place allouée par les usagers à l'actualité médiatique sur dans leur profil personnel, mais aussi dans la publicisation de leur identité numérique. Parce qu'elle portait sur des espaces numériques de nature distincte, cette observation ethnographique en ligne a, en outre, autorisé son auteure à effectuer une comparaison entre eux<sup>88</sup>.

Coralie Le Caroff mentionne également les difficultés engendrées par l'archivage des ressources documentaires au cours d'une observation ethnographique en ligne. Cet archivage demeure incontournable pour procéder à l'analyse des données. Toutefois, en raison de l'instabilité des médias socionumériques (leurs caractéristiques sociotechniques pouvant connaître des changements substantiels, et ce, sans qu'un avertissement préalable ne soit donné), du volume vertigineux des publications sur Facebook et des obstacles rencontrés pour accéder à d'anciennes publications, l'archivage se révèle être un problème de taille pour mener une observation ethnographique en ligne (Béliard et Le Caroff, 2018). Pour sa part, l'auteure a mis en place un procédé manuel de sauvegarde, reposant sur l'usage des captures d'écran.

---

87. En 2011, Coralie Le Caroff procède à l'observation de fils de discussion sur la mort de Kadhaï, les élections en Tunisie et la Charia en Libye sur les pages Facebook de BFM TV, de Courrier international, de l'Express, du Monde et du Figaro. En 2013, elle réitère cette opération en suivant les fils de discussion traitant du racisme en France, de l'augmentation du revenu de solidarité active (RSA), de l'augmentation du chômage, de l'avortement et du repas scolaire unique en France sur les pages Facebook du Monde, de BFM TV et du Figaro.

88. La grille d'observation n'est pas explicitement mentionnée dans le chapitre méthodologique de la thèse. Toutefois, dans un chapitre cosigné avec Josiane Jouet, Coralie Le Caroff expose clairement une grille adaptée à l'observation des usages des dispositifs sociotechniques, laquelle se trouve ensuite illustrée par une étude de cas sur les débats en ligne autour de l'émission *Le jeu de la mort*. La lecture de ce chapitre nous apparaît par là même profitable pour ceux qui souhaiteraient avoir un exemple de grille d'analyse sous-jacente à une observation ethnographique en ligne (Jouet et Le Caroff, 2016).

Le second volet de la méthodologie consiste en la réalisation de 35 entretiens semi-directifs et d'un entretien collectif (*focus group*) à la fin de son enquête. Les entretiens semi-directifs ont été effectués auprès d'usagers de Facebook correspondant à deux profils différents, les « partageurs » (ceux qui ne font que relayer des contenus médiatiques à leur réseau interpersonnel) et les « commentateurs » (qui, en plus de partager fréquemment des contenus médiatiques, s'expriment à leur sujet). Les personnes interrogées s'identifient, pour une partie d'entre elles, aux propriétaires des profils personnels ayant été observés. Les thèmes abordés à l'occasion de ces entretiens semi-directifs concernaient les « usages privés de Facebook », « les commentaires d'actualités sur Facebook », « l'expression politique sur Facebook », « l'expression politique en ligne hors Facebook », « l'expression politique en face à face » et « la politisation/l'engagement » (Le Caroff, 2015, p. 572-573).

En se référant aux travaux de Jean-Claude Kaufmann (1996), Coralie Le Caroff conçoit ses échanges avec les enquêtés sur le modèle de l'entretien compréhensif. Cela signifie ainsi d'abandonner une prétendue posture de neutralité dans les entretiens pour se montrer empathique, sinon sympathique, à l'égard des enquêtés. L'entretien compréhensif se situe donc en rupture avec l'idée d'un entretien entre quelqu'un qui sait et quelqu'un qui ignore, entre quelqu'un qui sait que l'autre ignore et quelqu'un qui ignore qu'il ignore. Cette approche devait ainsi permettre de surmonter l'asymétrie des rapports entre l'enquêteur et l'enquêté dans la réalisation des entretiens. Par ailleurs, cela confirme encore une fois les présupposés épistémologiques de la recherche, tournés vers l'interprétation, celle-ci visant à la fois à décrire le contexte des pratiques sociales en menant une observation ethnographique, et à saisir les significations qu'attribuent les usagers à leurs propres activités sur Facebook.

*Les résultats obtenus.* La richesse de cette thèse de doctorat fait obstacle au projet de récapituler exhaustivement toutes ses contributions scientifiques. Nous pouvons toutefois en donner un aperçu, à partir de la conclusion de ce travail (Le Caroff, 2015, p. 525-538), en nous focalisant sur quatre points. Sur la base de son dispositif méthodologique, une ethnographie en ligne, cette recherche parvient à fragiliser, sinon à invalider,

l'idée communément répandue d'une apathie politique grandissante dans les démocraties libérales occidentales. Bien que le périmètre de cette enquête de terrain soit assez limité, il n'en reste pas moins que les données recueillies et analysées démontrent l'existence d'un intérêt pour le politique parmi des acteurs « profanes », n'ayant ni un profil militant ni une carrière de professionnel de la politique. Si les enquêtés semblent peu enclins à s'émouvoir de la politique institutionnelle, leurs discours et leurs interactions en ligne témoignent d'une réelle préoccupation pour les affaires publiques ou, conformément à l'étymologie grecque du mot « politique », pour la vie collective (Lalande, 1993, p. 785).

Le deuxième enseignement de cette thèse de doctorat concerne les pratiques informationnelles, celles-ci connaissant des évolutions significatives. Cette investigation révèle en effet un éclectisme dans les rapports individuels aux contenus médiatiques, ainsi qu'une tendance à délaisser les supports traditionnels (quotidiens imprimés, téléjournaux, etc.) au profit des plateformes numériques pour s'informer, échanger et commenter l'actualité médiatique. Le fil d'actualité de Facebook se situe bien souvent au cœur des pratiques informationnelles des enquêtés, ce dispositif pouvant même être dans certains cas une unique source d'information. Ces changements sur le plan des pratiques informationnelles contribuent de ce fait, pour reprendre une expression de l'auteure, à « un effacement des médias sur Facebook » (Le Caroff, 2015, p. 527).

En troisième lieu, cette thèse de doctorat détaille les modalités par lesquelles les enquêtés s'expriment au sujet de l'actualité médiatique sur le réseau de Mark Zuckerberg. Or, la comparaison dévoile des différences marquées entre la parole des enquêtés sur les pages publiques des médias, d'une part, et celle qui s'exprime sur les profils personnels, d'autre part. Sur les pages publiques des médias, les enquêtés se montrent plus émotifs, réagissant souvent négativement à l'actualité médiatique. En revanche, sur les profils personnels, les énonciations des usagers s'avèrent moins spontanées, démontrant une tendance à la rationalisation de la présentation et de l'expression de soi dans leurs réseaux interpersonnels. Pour terminer, cette recherche met en lumière une atténuation des distinctions de genre dans les prises de parole sur Facebook. Bien que les rapports de genre

persistent dans les modalités d'expression des enquêtés, les frontières entre ce qui est perçu comme féminin ou masculin perdent de leur évidence dans les interactions médiatisées des usagers de ce réseau social.

*Les limites du dispositif méthodologique.* Cette thèse de doctorat met en exergue autant les potentialités que les limites inhérentes à la réalisation d'une ethnographie en ligne. À la lumière des résultats obtenus, cette stratégie méthodologique apparaît particulièrement efficace pour restituer avec finesse les pratiques informationnelles des enquêtés et accéder aux significations qu'attribuent ces acteurs à leurs usages sociopolitiques. La combinaison de l'observation ethnographique en ligne avec la conduite d'entretiens semi-directifs offre, par conséquent, la possibilité de décrire et de comprendre l'évolution des manières de s'informer aujourd'hui. On peut s'interroger bien sûr sur le degré de généralisation des conclusions d'une telle étude ethnographique. Il ne s'agit pas, en disant cela, de contester la validité de cette recherche qualitative, mais plutôt d'en questionner la portée. Cette limitation, nous le savons bien, concerne toutes les enquêtes de type ethnographique (Hammersley, 1992, p. 85-95). Tout se passe comme si la profondeur de l'analyse se devait d'être inversement proportionnelle au volume des données collectées. Une telle précision dans la description et la compréhension des pratiques informationnelles semble difficilement conciliable avec un élargissement du corpus. Voilà pourquoi plusieurs chercheurs de notre panel nous disaient privilégier l'option d'une hybridation des techniques d'enquête qualitatives et quantitatives<sup>89</sup>. En autorisant la construction d'un corpus plus vaste, les méthodes composites rendraient possible une généralisation des résultats, sans pour autant compromettre le pouvoir d'intelligibilité d'une approche qualitative.

En outre, Coralie Le Caroff reconnaît avoir eu des difficultés à recruter des participants dans le cadre de son observation des profils personnels sur Facebook et de ses entretiens semi-directifs<sup>90</sup>. Bien que ces complications ne soient pas intrinsèques

---

89. Voir le chapitre précédent du présent rapport.

90. Coralie Le Caroff fournit un compte rendu détaillé des difficultés de recrutement qu'elle a rencontrées (2015, p. 129-134).

à l'ethnographie en ligne, elles ont engendré un défaut de représentativité au niveau de la constitution de l'échantillon de cette recherche. Pour trouver des individus souhaitant contribuer à sa recherche scientifique, Coralie Le Caroff a été amenée à effectuer un « recrutement de proche en proche » (Le Caroff, 2015, p. 127). Cette opération s'est avérée efficace, mais malheureusement coûteuse quant à la diversité de son échantillon. Ce dernier témoignait en effet d'une relative homogénéité sociale. Les personnes ayant accepté de participer à l'étude jouissaient, dans leur grande majorité, d'un capital culturel non négligeable et appartenaient « à la petite bourgeoisie nouvelle » (Le Caroff, 2015, p. 124). Dès lors, les résultats auxquels mène cette ethnographie en ligne ne paraissent pas en mesure de refléter les pratiques informationnelles de l'ensemble de la population. En raison de ces biais dans la formation de son échantillon, le champ de validité des conclusions de cette étude se voit restreint à une fraction des usagers des réseaux sociaux.

### 3. Une étude sur le partage d'information par Twitter

Pour notre quatrième étude de cas, nous souhaitons examiner une recherche, menée par Axel Bruns, Tim Highfield et Stephen Harrington, sur le partage de l'information d'actualité par Twitter<sup>91</sup> intitulé « Sharing the News: Dissemination of Links to Australian News Sites on Twitter » (2013). Ce travail d'enquête s'intéresse non pas au partage d'information dans des situations exceptionnelles, à l'instar de conflits politiques majeurs, les catastrophes naturelles ou les événements médiatiques, mais à la circulation de contenus médiatiques entre les usagers de Twitter dans la routine de leur vie quotidienne. Elle examine, autrement dit, l'ordinaire des utilisations de cette plateforme en ligne. Loin d'être une initiative isolée, cette recherche s'inscrit en réalité dans un programme de recherche sur les usages et les

---

91. Bien que cette recherche repose spécifiquement sur l'application de techniques computationnelles aux traces des *big data*, en l'occurrence les données de Twitter, les auteurs de cette étude ne se font aucunement les apôtres de cette démarche méthodologique. Par exemple, Jean Burgess et Axel Bruns ont argumenté ailleurs en faveur d'une approche hybride, associant ces techniques computationnelles à d'autres techniques d'enquête quantitative (2012).

impacts des réseaux sociaux dans la société australienne. Ce programme articule plusieurs projets scientifiques, lesquels sont aisément accessibles sur la plateforme en ligne qui lui est consacrée (<http://mappingonlinepublics.net/>)<sup>92</sup>. S'il ne se limite pas à Twitter, ce programme de recherche accorde une attention particulière à ce média numérique en tant qu'objet d'étude, mais aussi en tant que ressource méthodologique<sup>93</sup>. Il rassemble en effet des travaux sur le rôle de Twitter dans le contexte d'une communication de crise liée à des inondations en Australie (Bruns, Burgess, Crawford et Shaw, 2012), dans le contexte d'une campagne électorale australienne (Bruns et Burgess, 2011) ou dans le cadre de la communication publicitaire d'Adidas (Krüger, Stieglitz et Potthoff, 2012).

*Pré-supposés épistémologiques.* Cette étude sur le partage de l'information d'actualité dans la « twittosphère » ne fait pas explicitement mention de ses orientations épistémologiques. Plusieurs indices tendent toutefois à pointer des affinités entre cette recherche et le positivisme logique (Carnap, 1973), posture que nous avons déjà esquissée plus haut. Axel Bruns, Tim Highfield et Stephen Harrington emploient en effet à maintes reprises les termes « observation » ou « observé » pour rendre compte de leur démarche scientifique. Ils formulent ainsi des énoncés d'observation ayant pour ambition de correspondre à des faits, en l'occurrence des récurrences (*patterns*) dans les activités des usagers de Twitter. Ces énoncés d'observation découlent d'un raisonnement inductif, lequel consiste formellement à respecter le principe qui suit : « si un grand nombre de A ont été observés dans des circonstances très variées, et si l'on observe que tous les A sans exception possèdent la propriété B, alors tous les A ont la propriété B » (Chalmers, 1990, p. 27). Le repérage de régularités dans les comportements des

---

92. Plusieurs publications discutent explicitement des aspects méthodologiques de ce programme de recherche. La question de la mobilisation des techniques computationnelles pour extraire et analyser les traces de Twitter fait même l'objet d'un traitement de faveur dans plusieurs articles. Nous utiliserons donc ces ressources bibliographiques pour cette sous-section du rapport (Burgess et Bruns, 2012; Bruns et Burgess, 2012a, 2012b; Bruns et Liang, 2012; Bruns et Stieglitz, 2013).

93. Ce qui est l'une des caractéristiques de l'approche des *digital methods*: voir le glossaire à la fin du présent rapport.



utilisateurs de cette plateforme en ligne procède donc, dans le cadre de ce travail, de l'observation d'un nombre très important de Tweets. Les auteurs entreprennent également de croiser leurs observations avec d'autres enquêtes quantitatives sur Twitter, celles-ci pouvant être pilotées par les mêmes personnes ou de proches collaborateurs (Bruns et Stieglitz, 2013). Loin d'être anecdotique, cette démarche autorise les chercheurs, d'une part, à généraliser certaines de leurs conclusions sur les usages de Twitter à l'ensemble de la population australienne et, d'autre part, à soutenir cette montée en généralité au-delà des singularités du partage de l'information d'actualité pendant une période restreinte<sup>94</sup>.

La recherche d'Axel Bruns, Tim Highfield et Stephen Harrington entretient une autre familiarité avec le positivisme logique, en ceci qu'elle adopte une conception probabiliste de ce qui fait la validité des connaissances scientifiques. Au regard du positivisme logique, le raisonnement inductif n'offre pas la possibilité de produire des propositions dont la validité serait vérifiable définitivement. Dans la mesure où chaque nouvelle observation demeure susceptible de corroborer ou d'infirmer les observations antérieures, le raisonnement inductif ne permet logiquement que de produire des connaissances scientifiques probables. Les auteurs, par la prudence de leurs analyses, semblent souscrire à une telle conception du savoir scientifique. Ils préfèrent en effet utiliser le vocabulaire de la probabilité plutôt que celui de la certitude quand ils formulent des propositions au sujet du partage de l'information de l'actualité par Twitter (« such patterns are likely », « it appears likely that those who most frequently », etc.). Ces deux dimensions – le raisonnement inductif et le probabilisme – concourent à laisser penser à une adhésion implicite de cette étude au positivisme logique.

---

94. Cette volonté de produire des connaissances générales – valant au-delà des singularités d'un cas particulier – sur les usages de Twitter explique pourquoi Axel Bruns et Stefan Stieglitz plaident en la faveur d'une standardisation des métriques de l'activité des utilisateurs de cette plateforme en ligne (2013). La standardisation des instruments de mesure leur apparaît nécessaire pour établir des comparaisons entre les différentes études sur Twitter et, subséquemment, pour construire des connaissances ayant une valeur générale.

*Assises théoriques.* Aucune théorie systématique ne vient soutenir cette recherche. Cette absence, loin d'être une source d'étonnement, demeure tout à fait cohérente d'un point de vue épistémologique. Parce que cette étude repose ultimement sur un raisonnement inductif, aucune ressource conceptuelle ne semble devoir être exigée pour émettre des énoncés d'observation. Tout se passe comme si, dans cette perspective, l'activité savante d'observation pouvait faire l'économie d'assises théoriques<sup>95</sup>. Ceci étant dit, les auteurs font une entorse à cette démarche inductive en empruntant la notion de *gatewatching* à d'autres publications d'Axel Bruns (2005). Si le *gatekeeping* fait référence aux opérations par lesquelles les journalistes effectuent une sélection – à partir de critères subjectifs et professionnels – parmi les informations auxquelles ils sont exposés (White, 2018), le *gatewatching* renvoie quant à lui aux processus par lesquels les usagers des médias numériques participent de manière collaborative à l'identification et à la circulation d'informations d'actualité jugées pertinentes. Cette notion permet à Axel Bruns, Tim Highfield et Stephen Harrington de circonscrire leur objet de recherche, tout en lui donnant simultanément une légitimité (2013, p. 1). Étudier les processus de *gatewatching* par la médiation de Twitter, c'est analyser un phénomène relativement inédit, en rupture avec les pratiques traditionnelles de *gatekeeping*. En outre, les auteurs mobilisent *a priori* une distinction entre la consommation (passive) des informations d'actualité et l'engagement (actif) que les usagers peuvent entretenir par rapport à l'actualité médiatique. Ce faisant, ils partent du présupposé que les pratiques informationnelles sur Twitter relèveraient davantage de la seconde catégorie (Bruns, Highfield

---

95. Cette idée d'une indépendance de l'observation par rapport à la théorie, que d'aucuns associent à « l'inductivisme naïf », reste extrêmement problématique. Il s'agit d'un problème classique en épistémologie, ayant donné naissance à des objections très sérieuses à l'endroit de cette représentation du travail d'observation (Chalmers, 1990, p. 38-72).

et Harrington, 2013, p. 2)<sup>96</sup>. Ils reprennent enfin, sans la remettre en question, la dichotomie information/opinion dans leur étude. Bien que celle-ci ne fasse nullement l'objet d'une explication, elle informe l'ensemble de leur démarche.

*Techniques d'analyse et de collecte de données.* Le dispositif méthodologique de cette étude repose sur la manipulation de données mises à disposition gratuitement par les propriétaires de Twitter. Pour accéder à ces données, les chercheurs ont exploité l'interface de programmation (*Application Programming Interface*) de cette plateforme en ligne. Contrairement à la technique du *scraping* qui ne requiert aucune entente préalable, ce procédé demeure conforme à la politique de l'entreprise et présente, de ce fait, l'avantage de préserver les scientifiques d'éventuelles poursuites judiciaires. Cette stratégie comporte toutefois de sérieuses limitations, dans la mesure où les chercheurs demeurent tributaires du libre arbitre – des intérêts commerciaux et financiers – de l'entreprise dans cette configuration<sup>97</sup>.

Si l'interface de programmation permet d'accéder aux données de Twitter, selon les modalités prévues par la compagnie californienne, les organisations désireuses de s'en servir doivent encore se doter d'une application appropriée. Axel Bruns, Tim Highfield et Stephen Harrington ont quant à eux opté pour le logiciel client (*open source*) *yourTwappperKeeper* (<https://github.com/540co/yourTwappperKeeper>), une évolution du désormais

---

96. Le concept d'engagement mériterait un approfondissement prenant appui notamment sur le travail théorique d'Howard S. Becker (« Notes on the Concept of Commitment », *The American Journal of Sociology*, 66 (1), 1960, p. 32-40). Selon Becker, ce concept n'exprime pas seulement, et de manière alors exclusivement descriptive, le fonctionnement d'un mécanisme psychosociologique de maintien, pour l'individu engagé, d'une ligne de « cohérence du comportement ». Par-delà cette perspective descriptive, le concept d'engagement peut aussi intégrer la notion de sens commun, ce qui permet alors d'expliquer pourquoi des « individus poursuivent souvent une ligne d'action pour des raisons tout à fait étrangères à l'action elle-même » (Becker). On voit que le concept d'engagement est sociologiquement plus complexe que la conception qui le fait équivaloir à une simple sollicitation répétée d'un même site web.

97. Nous reviendrons ci-dessous sur les risques que présente cette situation d'hétéronomie pour la recherche en sciences sociales.

obsolète *Twapperkeeper.com*<sup>98</sup>. Cet outil autorise la soumission de requêtes à l'interface de programmation de Twitter sur la base d'une combinaison de mots-clés et de hashtags<sup>99</sup>, l'exportation des résultats de ces requêtes sous différents formats (RSS, HTML, EXCEL et JSON) et, enfin, l'archivage de ces résultats. Les données recueillies par cette application portent, entre autres, sur le contenu des Tweets, les identifiants de leur envoyeur et de leurs destinataires, l'identifiant des Tweets, la langue utilisée, la date de publication des Tweets, etc. (Bruns et Liang, 2012). Ce logiciel offre également la possibilité d'exploiter deux fonctionnalités hautement pertinentes pour appréhender les pratiques informationnelles contemporaines. D'une part, il permet d'interroger l'interface de programmation de façon à collecter un ensemble de Tweets déjà publiés dans un passé plus ou moins proche. D'autre part, il permet aussi d'extraire en « temps réel » des Tweets qui viennent d'être publiés. Autrement dit, ces deux fonctionnalités donnent respectivement accès au passé ou au présent des manières de s'informer par Twitter. Dans le cadre de la recherche qui nous intéresse, c'est la seconde qui a été principalement mobilisée.

À l'aide de *yourTwapperKeeper*, les auteurs ont procédé à l'extraction en temps réel de Tweets contenant des liens URL pointant vers les sites web australiens de grandes organisations médiatiques (News Ltd, Fairfax, ABC, SBS, etc.) ou d'opinion (Crikey.com, independantaustralia.net, etc.). Ce travail a été effectué pendant deux mois, entre le 18 juin et le 19 août 2012.

---

98. L'évolution des conditions d'utilisation (*Terms of Service*) des données de Twitter s'avère être aussi rapide que déroutante pour les organisations tierces. La compagnie californienne modifie en effet fréquemment les règles entourant l'accès à ses données conformément à ses intérêts commerciaux et financiers. Ces changements, parfois brutaux, ne font d'ailleurs que peu l'objet d'une explicitation. Cette instabilité explique que des applications telles que *yourTwapperkeeper* semblent aujourd'hui – les développeurs de ce logiciel n'ont pas de certitude à ce sujet – contrevenir à la dernière version de ces conditions d'utilisation.

99. Axel Bruns et Jean Burgess définissent le « hashtag » de la manière suivante : « These brief keywords or abbreviations, prefixed by the hash symbol, are included in tweets – for example, #auspol for discussing of Australian politics, or #NotW to track the News of the World scandal. Hashtags are simply entered manually by users as they tweet; they can be created ad hoc, and emerge almost instantaneously as news breaks. » (2012a, p. 3).

Ils ont collecté pendant cette période plus d'un million de Tweets (1 318 928 Tweets au total). En procédant de la sorte, ils avaient à leur disposition les matériaux nécessaires pour quantifier ce que partagent les usagers de Twitter au regard de l'actualité médiatique et pour définir les caractéristiques des auditoires utilisant ce média numérique. Au total, 28 noms de domaines ont été traqués par les chercheurs, 21 en lien avec les sites web d'organisations médiatiques et 7 avec des sites web d'opinion (Bruns, Highfield et Harrington, 2013, p. 3-4)<sup>100</sup>. Les auteurs estiment, dans leur publication, avoir intégré l'essentiel des sites web consacrés à l'actualité médiatique en Australie. Ils entendaient ainsi produire des analyses représentatives des pratiques informationnelles des Australiens sur Twitter (p. 3). Il ne s'agit aucunement d'une prétention déraisonnable dans le contexte australien, dans la mesure où la population n'est pas dispersée sur l'ensemble du territoire (ce qui n'incite donc pas à la multiplication des organes de presse régionaux), d'une part, et où la structure de l'industrie médiatique demeure très concentrée dans cet État, d'autre part. Ce dispositif méthodologique sert, par ailleurs, des ambitions dépassant le cadre particulier de cette étude sur le partage de l'information d'actualité. Il a en effet donné naissance à un indice : *l'Australian Twitter News Index (ATNIX)*<sup>101</sup>.

Bien que ces outils ne soient pas mentionnés dans leur article, Axel Bruns, Tim Highfield et Stephen Harrington ont très probablement eu recours à *Gawk* (<http://www.gnu.org/software/gawk/>), un logiciel de manipulation de données, et *Gephi* (<https://gephi.org/>), un logiciel de visualisation de données, pour mener leur étude. Ces deux applications *open source* ont, à vrai dire, été mobilisées dans plusieurs recherches du programme *Mapping Online Publics*. Ces précisions concernant les outils étant

---

100. Chaque site web possède un nom de domaine. Les URL correspondent, quant à eux, aux adresses pointant vers des pages précises de chaque site web.

101. Cet indice, développé d'abord dans le contexte du programme de recherche *Mapping Online Publics*, est encore utilisé aujourd'hui pour suivre l'activité des usagers de Twitter en lien avec l'actualité médiatique (<http://mappingonline-publics.net/tag/atnix/>). Il se veut à la fois un instrument de mesure standardisé destiné à la communauté scientifique, avec pour objectif de faciliter le travail de comparaison entre les différentes études menées sur Twitter, mais aussi un outil pour les professionnels (Bruns et Stieglitz, 2013, p. 14)

faites, on note que les réflexions des chercheurs, dont nous exposerons les résultats plus loin, croisent trois paramètres. Le premier a trait à l'unité de temps de l'analyse, les auteurs examinant leurs données sur une base hebdomadaire ou quotidienne. Le second concerne la division entre ce qui révèle de l'information et ce qui touche l'opinion. Cette distinction traverse la totalité des analyses, pour déterminer autant les sources de l'actualité médiatique, les caractéristiques des auditoires sur Twitter que les contenus des Tweets partagés. Le troisième découle de la règle des 1/9/90 (Tedjamulia, Dean, Olsen et Albrecht, 2005), cette règle visant à différencier le degré d'engagement des usagers par rapport à l'actualité médiatique : 1 % seraient considérés comme étant les plus actifs, 9 % comme étant actifs et 90 % comme les moins actifs (Bruns, Highfield et Harrington, 2013, p. 10).

*Résultats obtenus.* Cette étude met en lumière plusieurs caractéristiques du partage de l'information d'actualité sur Twitter en Australie. Celles-ci sont récapitulées à la fin de l'article et peuvent être synthétisées en trois points. Les auteurs soulignent d'emblée une préférence chez les usagers pour les sources d'information australienne « de qualité », autant en ce qui a trait aux plateformes en ligne des organisations médiatiques que sur le plan des sites web d'opinion. Axel Bruns, Tim Highfield et Stephen Harrington identifient, sans s'expliquer d'ailleurs à ce sujet, ces sources d'information à des références sérieuses, intellectuelles et, en un certain sens, élitistes<sup>102</sup>. Ils les opposent explicitement aux « tabloïds populistes » (Bruns, Highfield et Harrington, 2013, p. 15), dont l'influence s'avère mineure dans la circulation des contenus d'information et d'opinion dans la « twittosphère » australienne. Ce premier résultat vient par conséquent fragiliser la thèse, quelquefois défendue dans la littérature scientifique, selon laquelle les manières de s'informer par les médias numériques témoigneraient d'une emprise grandissante de contenus ne répondant pas nécessairement aux exigences professionnelles du journalisme. Cette thèse semble devoir être nuancée au moins dans le contexte australien.

---

102. Les auteurs utilisent fréquemment l'expression « highbrow » pour qualifier ces sources d'information dans leur argumentation. Celle-ci connote des pratiques intellectuelles et élitistes, renvoyant parfois à la « haute culture ».

La deuxième conclusion consiste à établir une corrélation entre ce portrait des sources d'information et les propriétés sociodémographiques des usagers australiens de Twitter dans leur ensemble. Ceux-ci développeraient une appétence pour les contenus rigoureux d'information et d'opinion en raison de leur position dans l'espace social (il s'agirait en majorité d'individus âgés de 25 à 55 ans, habitant dans des milieux urbains et jouissant d'un niveau élevé de scolarité). Cette corrélation ne peut cependant pas être établie avec certitude, dans la mesure où, au moment de réaliser cette étude, les auteurs ne disposaient pas de données fiables et exhaustives sur le profil des utilisateurs de cette plateforme en ligne en Australie. La validité de ce résultat n'est donc que probable.

Enfin, la dernière conclusion concerne la répartition des pages web d'information par rapport aux pages web d'opinion. Il appert, en l'occurrence, que les Australiens font circuler plus de contenus en provenance de pages web d'information – celles-ci étant le plus souvent intégrées aux sites web des organisations médiatiques – que de contenus issus de pages web d'opinion dans leurs Tweets. L'auditoire des contenus d'opinion se révèle ainsi être quantitativement moins important que celui des contenus d'information dans la « twittosphère » australienne<sup>103</sup>.

*Limites du dispositif méthodologique.* Cette recherche présente l'avantage d'offrir des mesures sophistiquées et précises du partage de l'information d'actualité par Twitter<sup>104</sup>. En exploitant un corpus pour le moins imposant de traces (1 318 928 Tweets)

---

103. Les auteurs rappellent toutefois que leur étude ne donne pas une vision exhaustive des contenus d'actualité auxquels s'exposent les Australiens, puisque ces derniers consultent également des informations et des opinions en provenance de sites web internationaux. Leur recherche n'intègre que les sites web australiens.

104. Axel Bruns, Tim Highfield et Stephen Harrington voient dans leur dispositif méthodologique un moyen d'échapper aux limites des mesures d'auditoire traditionnelles, lesquelles ne permettraient pas de quantifier l'auditoire réel d'un contenu médiatique (une émission de télévision peut, par exemple, être retransmise sans qu'aucun téléspectateur n'y prête attention). Ces auteurs entendent capturer l'auditoire effectif des sites web australiens d'information et d'opinion en se concentrant uniquement sur des pratiques informationnelles requérant une activité minimale des usagers: le partage de Tweets (Bruns, Highfield et Harrington, 2013, p. 2).

à l'aide de techniques computationnelles de collecte, d'analyse et de manipulation des données, les auteurs décrivent plusieurs caractéristiques saillantes des pratiques informationnelles contemporaines en Australie. Bien qu'ils recourent, dans cette étude, à leur indice ATNIX pour suivre les manières de s'informer pendant une courte période, cet instrument de mesure apparaît également prometteur pour décrire des tendances beaucoup plus longues (Bruns, Highfield et Harrington, 2013, p. 2). Ce dispositif méthodologique ne manque donc pas d'attraits pour ceux qui souhaiteraient objectiver des phénomènes sur la base d'une quantification à grande échelle. Plusieurs limites – ayant des liens avec des considérations politiques, économiques et épistémologiques – viennent cependant modérer l'enthousiasme qu'il peut légitimement susciter.

Axel Bruns se montre d'abord très préoccupé par les contraintes – effectives ou possibles – engendrées par la situation d'hétéronomie dans laquelle se trouvent les recherches en sciences sociales empruntant cette voie méthodologique (Bruns et Burgess, 2012b, p. 10; Burgess et Bruns, 2012; Bruns et Liang, 2012). Ces enquêtes dépendent en effet directement de l'accessibilité des données des plateformes en ligne, celle-ci étant déterminée en fonction des stratégies commerciales ou financières de leurs propriétaires. Cela signifie que les règles relatives à l'accessibilité des données d'une entreprise peuvent, en évoluant, faire obstacle à la réalisation d'une étude, voire même forcer son interruption. Cette difficulté s'avère d'ailleurs particulièrement vive dans le cas de Twitter, cette compagnie adoptant depuis quelques années des règles de plus en plus restrictives, peu compatibles avec la félicité d'un projet de recherche.

De plus, ce dispositif méthodologique, en raison de son degré de sophistication technique, n'est pas à la portée de toutes les bourses. Il requiert, à vrai dire, des ressources financières conséquentes, puisqu'il faut enrôler des personnes dotées de compétences en ingénierie informatique et former dans ce domaine des chercheurs en sciences sociales, mais il faut aussi se procurer l'équipement adéquat pour extraire, analyser et produire des visualisations des traces des *big data*. Plus encore, certaines compagnies – Twitter, en tant qu'entreprise, appartient à cette catégorie – entendent tirer profit de la vente de leurs données à des tiers, et ce parfois à des prix prohibitifs. Ce coût constitue un obstacle



économique supplémentaire pour les chercheurs désireux de recourir aux *big methods*. Dans le cadre du programme *Mapping Online Publics*, plusieurs applications *open source* ont été mobilisées (*yourTwapperKeeper*, *Gawk*, *Gephi*, etc.) au détriment de logiciels propriétaires. Ces choix techniques ont pu contribuer à réduire les coûts d'implémentation de cette stratégie méthodologique. Il existe donc des tactiques permettant de réduire les dépenses allouées à une telle méthode. Il n'en reste pas moins que ce programme de recherche a été réalisé dans le contexte du Center of Excellence for Creative Industries (CCI), lui-même soutenu financièrement par l'Australian Research Council (ARC) pendant huit ans. Difficile, dans ces conditions, d'imaginer une étude reposant sur l'exploitation des données de type *big data* par des techniques computationnelles sans des ressources économiques importantes<sup>105</sup>.

Outre le problème de ses limites stratégiques et économiques, cette méthode n'offre pas la possibilité de comprendre en profondeur le sens social des pratiques informationnelles, c'est-à-dire de saisir les significations attribuées par les Australiens à leurs pratiques de partage de l'information d'actualité, d'une part, et d'approfondir le contexte dans lequel celles-ci se déploient, d'autre part. Ayant déjà abordé ce point précédemment, nous ne souhaitons pas revenir sur la question de la compréhension. Notons néanmoins que les chercheurs du programme *Mapping Online Publics*, loin d'être aveugles à cette insuffisance, défendent l'idée selon laquelle il faudrait associer les *big methods* à des méthodes qualitatives pour appréhender pleinement les phénomènes communicationnels (Burgess et Bruns, 2012). Ils plaident notamment en faveur d'une combinaison/hybridation entre les techniques de collecte et d'analyse décrites ci-dessus et des études de cas de type ethnographique, pour surmonter les limites des méthodes strictement computationnelles et quantitatives.

---

105. L'un de nos enquêtés, chercheur ayant des compétences à la fois en sociologie et en informatique, prétend pouvoir réaliser un tel type de recherche à un coût moindre en utilisant le serveur de son institution universitaire et en engageant une assistante de recherche qualifiée pendant un an.

#### 4. Une étude quali-quantitative sur les usages de Facebook

Pour terminer ce chapitre, nous prendrons comme étude de cas la thèse de doctorat d'Irène Bastard intitulée « De proches en pages, de pages en proches. Exploration et réception des informations en ligne » (2015). Deux articles, dont Irène Bastard est l'auteure principale, nous aideront également à rendre compte de sa démarche, l'un publié dans la revue *Sociologie* (2017) et l'autre sur le site web *InternetActu.net* (2013). Cette recherche doctorale mobilise une méthode numérique dite « quali-quantitative ». Celle-ci est aux méthodes numériques ce que le genre est à l'espèce : tout comme les méthodes numériques de la *Digital Methods Initiative* (DMI)<sup>106</sup>, il s'agit d'une forme particulière de *digital methods*. Rappelons que ces dernières, au sens général du terme, partagent au moins trois caractéristiques communes. Elles sont conçues : pour s'appliquer aux recherches en sciences sociales ; pour que l'on puisse tirer profit des objets nativement numériques ; et, enfin, pour permettre l'articulation de techniques d'enquête quantitatives et qualitatives.

Cette thèse de doctorat entend proposer « une description des pratiques du partage d'information et un questionnement sur ses implications [dont] l'enjeu est de comprendre les mécanismes sociaux qui mettent en circulation et en réception les informations sur le web » (2015, p. 35). La plateforme de Mark Zuckerberg se trouve, en l'occurrence, au cœur de cette étude. La question « qui partage quoi avec qui ? », faisant d'ailleurs clin d'œil (en forme de variation) au questionnement programmatique classique d'Harold Lasswell (qui ? dit quoi ? à qui ? par quel canal ? avec quels effets ?), se trouve au principe de l'ensemble de la démarche (2015, p. 40). Cette interrogation permet de cerner rapidement les trois grands axes d'analyse de cette thèse de doctorat. Le « qui ? » renvoie aux propriétés sociodémographiques des acteurs faisant circuler de l'information en ligne, mais aussi aux compétences exigées par cette activité (2015, p. 41). Le « quoi ? » concerne la nature des contenus d'actualité

---

106. <http://asca.uva.nl/content/research-groups/digital-methods-initiative/digital-methods-initiative.html>

faisant l'objet d'un partage. Il s'agit notamment de déterminer leur degré de « confidentialité », mais aussi si ceux-ci relèvent de l'information ou de l'opinion (2015, p. 42). Le « avec qui? » se rapporte au type de liens sociaux en jeu dans ces activités de partage. La distinction entre les liens faibles et les liens forts<sup>107</sup> vient structurer cette réflexion sur les propriétés des relations sociales sous-jacentes à la circulation des contenus d'actualité par Facebook (2015, p. 42-43).

*Présupposés épistémologiques.* À l'instar des autres recherches examinées précédemment, la thèse de doctorat d'Irène Bastard n'explique pas son positionnement épistémologique. Ceci étant dit, il est possible d'entrevoir ce dernier en prêtant attention à l'architecture méthodologique de l'étude. Ce travail repose en effet sur une association entre des techniques d'enquête « conventionnelles » (qualitatives et quantitatives) et des techniques computationnelles exploitant les *big data*. L'auteure conjugue, ce faisant, deux modes d'intelligibilité dans ses analyses, à savoir la compréhension, d'une part, et la description par la mesure, d'autre part<sup>108</sup>. Ce point ne fait d'ailleurs aucun doute à la lecture de l'introduction de la thèse (2015, p. 45-48). Dans cette dernière, Irène Bastard défend d'emblée l'idée selon laquelle « les usages du web ne peuvent s'envisager sans allier du quantitatif et du qualitatif, sans mesurer les activités et laisser les enquêtés dire le sens qu'ils donnent à cette activité » (2015, p. 45). Elle légitime cette association entre le quantitatif et le qualitatif en mettant en évidence une circularité dans la construction des connaissances sur les activités en ligne. D'après l'auteure, les mesures donnent à vrai dire une assise solide à la description, mais permettent aussi de préparer avec efficacité les entretiens en

---

107. « Les liens forts sont ceux que l'on a avec des amis proches (il s'agit de relations soutenues et fréquentes). Les liens faibles sont faits de simples connaissances. Les liens faibles sont dits « forts » dans la mesure où, s'ils sont diversifiés, ils permettent de pénétrer d'autres réseaux sociaux que ceux constitués par les liens forts. Les liens absents sont ceux caractérisés par une absence d'interaction. D'après Mark Granovetter (*Strength of weak ties*, 1973), cela peut être le marchand de journaux que l'on croise tous les matins ou un voisin dont on connaît seulement le nom. » (Source : [https://fr.wikipedia.org/wiki/Mark\\_Granovetter](https://fr.wikipedia.org/wiki/Mark_Granovetter))

108. Comme nous avons déjà abordé ces concepts, centraux pour caractériser épistémologiquement les sciences empiriques, nous renvoyons nos lecteurs à nos développements antérieurs pour avoir des éclaircissements à ce sujet.

accédant à des informations sur les pratiques des acteurs (2015, p. 48). À l'inverse, elle voit dans les entretiens une technique d'enquête indispensable pour appréhender les significations que les acteurs attribuent à leurs pratiques et, ce faisant, pour donner du sens aux mesures de leurs traces sur Facebook. Aussi, cette rencontre du quantitatif et du qualitatif est plus sophistiquée qu'une simple juxtaposition de deux stratégies méthodologiques. Cette thèse de doctorat semble ainsi pouvoir être interprétée comme une tentative de démontrer « sur pièce » la fécondité d'un dispositif méthodologique reposant sur cette circularité entre le qualitatif et le quantitatif.

*Assises théoriques.* Cette thèse de doctorat ne repose pas sur une théorie unifiée et systématique. Elle se fonde plutôt sur un ensemble de notions, parfois composites, répondant en premier lieu à une exigence de souplesse. L'auteure explique en effet vouloir des notions suffisamment ouvertes et flexibles pour être en mesure de s'adapter aux données qui ont émergé au cours de son travail de terrain. Dans cette perspective, les notions guident la recherche, sans toutefois l'enfermer dans un cadre conceptuel préétabli. Trois notions se trouvaient en l'occurrence au cœur de la construction de la problématique de cette recherche : l'information, l'activité en ligne et le partage.

L'information fait l'objet d'une double définition. Cette dualité, bien qu'étonnante en première analyse, se comprend aisément au regard du caractère bicéphale du dispositif méthodologique de l'étude. Lorsqu'elle collecte et analyse des données qualitatives, Irène Bastard ne prétend pas déterminer *a priori* le sens de la notion d'information. Parce qu'elle s'intéresse à l'information du point de vue des publics, elle privilégie la définition qu'en donnent les enquêtés. Ce n'est donc pas l'information en soi qui suscite l'attention de l'auteure, mais plutôt ce que signifie l'« information » pour les publics. En revanche, dans le cadre du moment quantitatif de sa recherche, l'information reste identifiée « aux contenus produits par les médias » (2015, p. 26-27).

La notion « d'activité en ligne » comporte quant à elle trois dimensions. La première concerne les pratiques médiatiques en ligne, que l'auteure envisage par rapport à la position des acteurs dans l'espace social. Plus encore, ces pratiques médiatiques en

ligne semblent pouvoir être distinguées en fonction de deux groupes d'individus : les consommateurs réguliers d'information et les consommateurs opportunistes (2015, p. 27-28). Dans la mesure où ces pratiques s'inscrivent dans le cadre de leurs relations sociales, la sociabilité en ligne forme la deuxième dimension de la notion d'activité en ligne. Cette dernière est appréhendée à l'aide de la microsociologie goffmanienne (2015, p. 29). La troisième dimension a trait à la manière dont les plateformes en ligne affectent les réseaux sociaux (au sens d'ensembles de liens sociaux).

La notion de « partage de l'information » vient terminer ce triptyque. Elle est définie comme « une activité interactionnelle qui utilise un contenu informationnel » (2015, p. 32). L'auteure insiste, entre autres, sur le double visage des acteurs participant au partage de l'information en ligne. Ceux-ci endossent en effet autant le rôle d'émetteur que celui de récepteur dans ce processus, ce qui implique d'abandonner les modèles traditionnels de la communication, reposant sur une distinction claire entre ces deux figures (2015, p. 33).

*Techniques de collecte et d'analyse des données.* Le dispositif méthodologique de cette thèse de doctorat s'avère éminemment complexe par rapport aux enquêtes examinées précédemment. Ce travail, parce qu'il entend tirer profit de la circularité entre les techniques d'enquête conventionnelles et les techniques computationnelles appliquées aux *big data*, ne peut que difficilement échapper à ce niveau de sophistication. Rappelons à ce propos que l'auteure adhère au principe selon lequel « sans cesse, dans le travail conduit avec les données du web, un aller-retour entre informations numériques et informations obtenues par les techniques d'enquête traditionnelles apparaît nécessaire » (Bastard, Cardon, Fouetillou, Prieur et Raux, 2013).

La stratégie méthodologique a été déterminée après une phase exploratoire, au cours de laquelle Irène Bastard a pu affiner ses techniques d'enquête et peser la pertinence de ses objets de recherche (2015, p. 45-46). Dans ce cadre, elle a d'abord effectué une analyse diachronique – entre 2002 et 2012 – de différentes technologies numériques autorisant le partage de l'information en ligne (fils RSS, réseaux socionumériques, etc.). Cette opération visait à documenter les idéologies sous-

jaçentes à la conception de ces technologies numériques, mais aussi à mettre en lumière l'évolution de ces mêmes idéologies. Cette étape exploratoire a également été à l'origine d'une évaluation de « la taille de l'activité de partage d'information en partant des contenus produits par les médias » (2015, p. 45). Il s'agissait de prêter attention au volume des commentaires publiés par les usagers autour de certaines actualités médiatiques sur Le Monde.fr et Rue89, mais aussi aux compteurs d'activités de Facebook relatifs à 18 000 articles de presse pendant les élections présidentielles françaises de 2012. Pour terminer, cette phase exploratoire a été l'occasion de réaliser 16 entretiens semi-qualitatifs avec un panel d'usagers relativement peu diversifiés d'un point de vue sociodémographique. Cette procédure a permis d'expérimenter et d'améliorer les grilles d'entretien qui ont été utilisées par la suite par Irène Bastard. La conjugaison de ces trois opérations a ultimement confirmé la pertinence d'une étude des pratiques de partage de l'information en ligne par Facebook (2015, p. 46).

À partir de ces résultats préliminaires, l'auteure a développé une stratégie méthodologique à deux volets. Le premier reposait sur des techniques d'enquête conventionnelles autant quantitatives que qualitatives. Dans le cadre de cette thèse de doctorat, l'auteure ne pouvait pas raisonnablement envisager d'étudier toutes les pratiques de partage de l'information sur Facebook. Pour d'évidentes raisons de faisabilité, Irène Bastard a choisi de restreindre le périmètre de sa recherche aux adolescents et, plus spécifiquement, aux lycéens<sup>109</sup>. Dès lors, elle a conçu et soumis un questionnaire à 562 lycéens, entre les mois de mars et juin 2013, dans l'idée d'en faire ensuite un traitement statistique. Cet échantillon, très équilibré sur le plan de la répartition des sexes,

---

109. La décision de se focaliser sur les adolescents n'était pas, loin de là, motivée uniquement par une exigence de faisabilité. Les adolescents sont en effet à une étape de leur vie où ils se construisent en fonction de leurs relations sociales et de leurs centres d'intérêt. De ce point de vue, leurs pratiques informationnelles sur le réseau Facebook engagent des enjeux identitaires. Ils possèdent également une certaine dextérité avec les technologies de communication numérique. Enfin, les pratiques sociales des adolescents semblent, comparativement à d'autres franges de la population, plus homogènes. Ces trois paramètres contribuent donc à rendre les adolescents attractifs pour étudier les pratiques informationnelles contemporaines (Bastard, 2015, p. 181-182).

n'était cependant pas représentatif de l'ensemble de la population lycéenne en France. La majorité des individus de l'échantillon provenait en effet des classes populaires (2015, p. 183-184). Dans ce questionnaire, l'auteure proposait notamment aux enquêtés une liste de sujets d'actualité susceptibles d'apparaître sur leur fil d'actualité Facebook. Les lycéens devaient alors se prononcer sur les sujets qu'ils auraient tendance à consulter, qu'ils commenteraient ou auxquels ils attribueraient un « *like* », mais aussi qu'ils seraient enclins à discuter sur le réseau (2015, p. 188-189). Ce questionnaire a également permis de recueillir des données sur le profil sociodémographique des enquêtés. Pour compléter cette première approche quantitative, Irène Bastard a réalisé 19 entretiens de groupe (*focus groups*) dans différentes classes, mais aussi 10 rencontres en tête à tête avec des lycéens (ceux-ci ayant été recrutés au cours des entretiens de groupe). Dans plus de la moitié des entretiens de groupe, les thèmes de l'amitié sur Facebook, de l'évitement de la politique et de la violence dans les interactions médiatisées ont été abordés (2015, p. 189). La ligne directrice de tous les entretiens de groupes demeurait toutefois la question des discussions autour de l'actualité sur Facebook (2015, p. 187). Les entretiens individuels étaient quant à eux l'occasion d'approfondir certains points avec des lycéens désireux de s'exprimer davantage sur la question de l'usage de Facebook.

Le deuxième volet de la stratégie méthodologique de cette thèse de doctorat s'inscrit dans le cadre du projet Algopol<sup>110</sup>. Or, l'une des particularités de ce projet réside dans l'ambition d'appliquer des techniques computationnelles aux traces des *big data*. Irène Bastard va d'ailleurs explicitement dans ce sens lorsqu'elle associe ce projet à « la catégorie des *very big data* en sciences sociales » (2015, p. 312). L'enjeu du projet Algopol consistait à « collecter de larges bases de données d'activité d'internautes (sur Twitter et les blogs initialement ; sur Facebook grâce à un protocole innovant), afin d'étudier les recommandations algorithmiques et sociales produisant la circulation des contenus en ligne » (2015, p. 47). Il a réuni pas moins d'une

---

110. Il est possible de prendre connaissance du projet Algopol en consultant le site web <http://algopol.huma-num.fr>.

trentaine de collaborateurs ayant des compétences en sociologie et/ou en informatique (<http://algopol.huma-num.fr/node/5>). Dominique Cardon (chercheur à Orange Labs, aujourd'hui professeur au Medialab à Science Po Paris), Irène Bastard (doctrante à Télécom ParisTech, aujourd'hui rattachée à la Bibliothèque nationale de France), Raphaël Charbey (Université Paris-Diderot), Jean-Philippe Cointet (CNRS), Baptiste Fontaine (Université Paris-Diderot), Christophe Prieur (Université Paris-Diderot, aujourd'hui à Télécom ParisTech) et Stéphane Raux (Linkfluence) ont joué un rôle de premier plan dans la conduite de ce projet.

Dans le contexte du projet Algopol, une application éponyme a été développée pour collecter et traiter un corpus volumineux de traces en provenance des utilisateurs de Facebook<sup>111</sup>. Ce logiciel ne manipule toutefois pas indistinctement toutes les traces des usagers. Il est en effet conçu pour extraire le profil, le mur et le réseau social (les « amis ») de leur compte Facebook. Irène Bastard a par conséquent adopté cette application pour le « versant computationnel » de sa méthode de recherche, avec pour objectif d'étudier les pratiques de partage de l'information sur ce réseau social. Loin d'être naïve, l'appropriation de ce logiciel a été menée en toute connaissance de cause. Les représentations sociales et les attentes normatives sous-jacentes à l'élaboration de ce logiciel sont en effet formulées explicitement par Irène Bastard (2015, p. 301-304). Cette application répond, en l'occurrence, à trois grandes attentes. Il y a d'abord une volonté d'étudier les activités en ligne à partir des individus et non des contenus médiatiques. Cette perspective requiert alors des chercheurs qu'ils obtiennent le consentement des usagers (conformément aux dispositions légales en France) et qu'ils s'assurent de la protection de leurs données personnelles. L'application Algopol répond ensuite au désir d'appréhender les activités en ligne à partir de données témoignant de l'activité effective des usagers en ligne et non de données déclaratives (récoltées *a posteriori* grâce à des questionnaires, d'entretiens,

---

111. L'application Algopol a été opérationnelle de décembre 2013 à avril 2015. Les évolutions successives de l'interface de programmation (*Application Programming Interface*) de Facebook ont fait progressivement apparaître des incompatibilités et, par suite, ont forcé l'interruption de son fonctionnement.



etc.). Irène Bastard parle même de la possibilité d'accéder « aux usages réels » – elle veut dire, sans doute, les usages « effectifs » – en utilisant, par l'application Algopol, l'interface de programmation de Facebook (2015, p. 303). Cette approche permettrait ainsi aux chercheurs de se prémunir contre les distorsions de la mémoire des enquêtés (ou les mésinterprétations de ces derniers par rapport aux questions posées par l'enquêteur). La conception de l'application Algopol part enfin du principe qu'il est pertinent de connaître les profils sociodémographiques des enquêtés pour analyser leurs activités en ligne, mais aussi de saisir les articulations entre leurs pratiques « en ligne » et leurs pratiques « hors ligne ».

Ces trois attentes prennent concrètement la forme d'un logiciel client-serveur, soumettant aux enquêtés une demande pour extraire leur profil, leur mur et leurs amis de leur compte Facebook. L'application Algopol offre, en échange de cette participation, la possibilité de visualiser, sous la forme d'une carte interactive dérivée de la théorie des graphes (2015, p. 298), le réseau amical des enquêtés sur cette plateforme en ligne. Les utilisateurs désireux de s'impliquer dans cette recherche doivent donc consentir à la manipulation de leurs données personnelles, mais aussi apporter, par l'intermédiaire d'un formulaire, quelques précisions sur leur situation (âge, sexe et profession) et leurs pratiques informationnelles. Contrairement à ce qui précède, l'application Algopol n'a pas été mobilisée pour étudier des comptes Facebook de lycéens, mais une pléthore de comptes dont les propriétaires possédaient des profils sociodémographiques variés (majoritairement des cadres, membres de professions libérales ou intellectuelles et des étudiants, mais aussi des retraités, des membres de professions intermédiaires, des ouvriers, etc.).

Les individus dont les comptes Facebook ont été manipulés par le logiciel Algopol ont été recrutés en deux phases distinctes, et ce sur la base d'une « auto-sélection » (Bastard, Cardon, Charbey, Cointet, Prieur, 2017, p. 61). La première phase a été réalisée dans le contexte d'une collaboration avec l'Institut CSA. Elle a été l'occasion d'étudier un échantillon de 831 enquêtés en novembre 2013. Au reste, la constitution de cet échantillon s'est vue soumise à une exigence de représentativité, de manière à refléter la réalité sociodémographique de la population

française. Si la méthodologie de l'Institut CSA garantissait une représentativité relativement contrôlée (l'échantillon ayant été formé en s'assurant de sa conformité aux quotas de représentativité des usagers d'internet en France), elle n'a cependant pas pu être aussi aboutie que pour un sondage classique (2017, p. 61). Pour la deuxième phase, un échantillon de 12 700 comptes a été réuni, en recrutant cette fois par « la communication autour de l'application, le bouche-à-oreille, et les réseaux de recherche et enseignement » (2015, p. 361). Cet échantillon ne pouvait pas toutefois prétendre être représentatif de la population française des internautes en raison des contingences du processus de recrutement des enquêtés<sup>112</sup>.

En ce qui concerne les techniques d'analyse de ce corpus volumineux de traces, Irène Bastard explique le caractère nécessaire, mais insuffisant, des techniques traditionnelles. Elle montre en effet, à partir des résultats de l'étude de l'Institut CSA, le peu de pertinence des indicateurs classiques, comme la moyenne, l'écart-type ou la médiane (2015, p. 317). L'insuffisance de ces derniers s'explique par le contexte singulier des activités en ligne. Elles évoluent de manière très rapide et discontinue, d'une part, et paraissent se conformer au modèle de la longue traîne<sup>113</sup>, d'autre part. Dans ces conditions, Irène Bastard a développé, avec les autres collaborateurs du projet Algopol, divers outils pour décrire les pratiques informationnelles des usagers de Facebook. Elle a notamment adopté cinq critères pour analyser les données de ce réseau socionumérique : « [i]l s'agit d'identifier (1) qui fait l'action, (2) quel objet utilise l'action, (3) où se fait l'action, (4) avec qui se fait l'action, (5) comment se fait l'action » (2015, p. 365).

*Résultats obtenus.* Il est impossible d'exposer, dans le cadre de cette sous-section, toutes les contributions de cette recherche à

---

112. Les propriétés des deux échantillons sont comparées et synthétisées dans le chapitre 7 de la thèse (2015, p. 363).

113. Appliquée en économie, l'expression « longue traîne » (*long tail*) est connue du grand public pour désigner la stratégie consistant à vendre une grande diversité de produits, chacun en petite quantité. Cette expression a été popularisée par Chris Anderson, en 2004, dans un article de la revue *Wired*, où il présente des entreprises telles qu'Amazon ou Netflix comme des exemples d'application de cette stratégie. (Source : [https://fr.wikipedia.org/wiki/Longue\\_traîne](https://fr.wikipedia.org/wiki/Longue_traîne)).

la connaissance des pratiques informationnelles. Le résumé que nous proposons ne rend donc pas justice à la richesse des apports de cette thèse de doctorat. Pour nous aider à les synthétiser, nous suivrons l'argumentation déployée par Irène Bastard dans la conclusion de son travail (2015, p. 389-393). Cette enquête s'efforce de répondre à la question « qui partage quoi avec qui ? » pour saisir les pratiques informationnelles contemporaines à partir de plusieurs points de vue, chacun d'entre eux étant élaboré dans le cadre de moments méthodologiques distincts (exploratoire, conventionnel et computationnel).

- « Qui ? » – Cette thèse de doctorat révèle d'abord que tous les usagers de Facebook participent au partage de l'information d'actualité à des degrés divers. Pour être plus précis, il se dégage deux grandes catégories de publics, l'un étant composé d'individus faisant circuler avec régularité l'actualité médiatique et l'autre d'usagers partageant l'information de manière occasionnelle (2015, p. 389). D'après les résultats de l'enquête, les individus les plus actifs par rapport à l'actualité médiatique sur Facebook demeurent clairement minoritaires. Contrairement à une représentation répandue, les jeunes usagers de Facebook n'utilisent pas intensément cette plateforme en ligne pour s'exprimer au sujet de l'actualité médiatique. Ils se servent surtout de celle-ci pour leur socialisation (2015, p. 389). Le partage de l'actualité médiatique n'est souvent qu'un moyen de nourrir leurs relations sociales. Il existe plusieurs sous-groupes dans la catégorie des individus actifs : « les semi-professionnels du partage de l'information » ; « les crieurs publics » ; les jeunes adultes inactifs, mais « partageurs intensifs » ; les technophiles (2015, p. 389-390).
- « À qui ? » – La zone de partage de l'information médiatique varie en fonction de l'interlocuteur ciblé. La variabilité des relations sociales influence donc autant les contenus que les modalités du partage de l'information. La recherche démontre que l'on « ne partage pas les mêmes mots avec ses proches, ses collègues ou des inconnus » (2015, p. 390).
- « De quoi parle-t-on ? » – Cette recherche confirme que les informations produites par les organisations médiatiques sont partagées par les usagers de Facebook (11 % des contenus partagés sur ce réseau social proviennent des

médias). Ceci étant dit, les contenus échangés par les internautes – que ceux-ci soient des adultes ou des adolescents – demeurent le plus souvent « familiers » (2015, p. 391). Les adolescents ont pour particularité de partager de l'information médiatique non pas en raison de l'intérêt qu'ils portent au contenu, mais d'abord en relation aux réactions que celui-ci peut engendrer chez le destinataire.

*Limites du dispositif méthodologique.* Les contributions de cette recherche sur les pratiques de partage de l'information sur Facebook s'avèrent extrêmement riches, tant sur le plan descriptif que d'un point de vue compréhensif. L'articulation des techniques d'enquête conventionnelles et des techniques computationnelles appliquées aux traces des *big data* ouvre, à l'évidence, des perspectives très prometteuses pour rendre compte des activités en ligne de manière multidimensionnelle. Cette approche n'en connaît pas moins certaines limites, à la fois techniques, matérielles et épistémiques.

Parce que cette méthode hybride repose sur des techniques computationnelles, elle rencontre les mêmes difficultés que les *big methods*. Comme nous l'avons vu à l'occasion de l'étude d'Axel Bruns, Tim Highfield et Stephen Harrington, les plateformes en ligne évoluent sans cesse, à un rythme souvent rapide et dans des directions qu'il est parfois difficile d'anticiper. Le succès des recherches demeure par conséquent tributaire des choix techniques des concepteurs de ces médias numériques, ces décisions étant elles-mêmes déterminées en fonction d'intérêts commerciaux et financiers. Cette instabilité des dispositifs sociotechniques s'avère donc un obstacle pour la recherche en sciences sociales, comme en témoigne, d'ailleurs, l'impossibilité de maintenir en service l'application Algopol à partir du mois d'avril 2015, à la suite d'un changement dans l'interface de programmation de Facebook. En outre, cette stratégie méthodologique hybride requiert d'importantes ressources matérielles, ce qui produit nécessairement un effet discriminant dans le monde de la recherche. Effectivement, toutes les organisations scientifiques ne disposent pas des moyens pour financer un dispositif méthodologique aussi sophistiqué, d'une part, ou pour réunir des équipes pluridisciplinaires composées de chercheurs en sciences sociales et d'informaticiens, d'autre part. Enfin, sur le plan épistémique, il importe de ne pas fétichiser le recours aux traces des

*big data*. Contre l'idée selon laquelle les données des plateformes en ligne permettraient de connaître les individus dans leur totalité, Irène Bastard rappelle que « [l']action, notamment en ligne, ne dit pas tout de l'acteur » (2015, p. 308). Les données collectées et analysées par les techniques computationnelles n'offrent toujours qu'un accès partiel aux comportements des acteurs. Il faut donc se garder de l'illusion d'une perspective totale sur les acteurs, qui permettrait d'appréhender leurs pratiques dans toutes leurs dimensions. De plus, l'auteure rend compte des difficultés liées à l'utilisation de ces corpus volumineux de traces. Loin d'être transparentes et utilisables immédiatement, les données des plateformes en ligne, pour être exploitables, exigent qu'un travail titanesque soit réalisé dans le cadre d'une recherche. Irène Bastard montre bien la nécessité de procéder au *reverse-engineering*<sup>114</sup> pour « faire parler les données » (2015, p. 312-319), et ce, même s'il est illusoire d'espérer supprimer toutes les ambiguïtés, les floues et les équivocités des données colligées par les médias numériques. Elle expose, par exemple, les difficultés suscitées par la manipulation d'une métadonnée centrale pour sa recherche sur les pratiques informationnelles, en l'occurrence le champ *status-type* indiqué par Facebook. Ce champ devrait permettre, en théorie, d'identifier les activités de chaque utilisateur de Facebook et d'en déterminer le volume. Or, il s'est avéré impossible de l'utiliser en l'état. Un travail considérable a donc été effectué par l'équipe d'Algopol pour donner du sens à cette métadonnée (Bastard, Cardon, Charbey, Cointet et Prieur, 2017, p. 63).

\* \* \*

Ce chapitre expose isolément les potentialités et les limites de chacun de nos idéaux-types de méthode de recherche en lien avec le numérique. Notre développement analytique s'attache, à partir de cinq études de cas sur les pratiques informationnelles contemporaines, à examiner pour chacune d'entre elles leurs présupposés épistémologiques, leurs fondements théoriques, leurs techniques de collecte et d'analyse de données, leurs pers-

---

114. Le *reverse-engineering* est l'activité qui consiste à étudier un objet pour en déterminer le fonctionnement interne ou la méthode de fabrication. (Source : <https://fr.wikipedia.org/wiki/Rétro-ingénierie>)

pectives de résultats et, enfin, leurs limites. Pour être en mesure de formuler des recommandations au Centre d'études sur les médias (CEM), il nous faut à présent passer à une dernière étape, à savoir une analyse comparative. C'est en procédant de la sorte que nous serons effectivement en mesure de déterminer le champ légitime d'application des méthodes conventionnelles qualitatives et quantitatives, de l'ethnographie en ligne, des techniques computationnelles appliquées aux *big data* et des méthodes numériques quali-quantitatives. Le prochain et dernier chapitre viendra clore ce rapport en s'attelant précisément à comparer ces approches méthodologiques et, sur cette base, à énoncer six recommandations au CEM.

# 5

## Conclusions

**P**our conclure ce rapport sur l'actualité des méthodes de recherche sur les pratiques informationnelles en sciences sociales, nous souhaitons procéder en trois temps. Nous résumerons d'abord les acquis les plus importants de notre parcours d'enquête. Puis, nous reviendrons sur les cinq recherches présentées dans le chapitre IV pour en proposer une analyse comparative. Il s'agira, ce faisant, de mettre en exergue les (dis) similitudes entre les méthodes sous-jacentes à chacune d'entre elles, mais aussi d'évaluer par effet de contraste les potentialités et les limites des différents idéaux-types de méthode de recherche en lien avec le numérique (les techniques d'enquête conventionnelles qualitatives et quantitatives, l'ethnographie en ligne, les techniques computationnelles appliquées aux traces des *big data* et, enfin, les méthodes numériques quali-quantitatives). Sur cette base, nous terminerons en formulant six recommandations au Centre d'études sur les médias concernant les stratégies méthodologiques qu'il pourrait être pertinent de mobiliser dans les études futures sur les pratiques informationnelles au Québec, au Canada et à l'international.

### **Quels sont les principaux points à retenir de ce rapport ?**

Le chapitre II met en lumière les problèmes et les enjeux des aspects méthodologiques de l'étude des pratiques informationnelles contemporaines. Pour ce faire, il s'appuie sur l'expertise

de 24 chercheurs francophones en sociologie, en communication et en gestion. Ayant tous été impliqués dans la conception de dispositifs méthodologiques pour collecter et analyser des données sur des activités en ligne, leur parole s'est avérée précieuse pour nous permettre de saisir les principales coordonnées des débats scientifiques autour des méthodes de recherche sur les manières de s'informer. Trois dimensions ont été abordées de façon récurrente par les personnes que nous avons interrogées : le diagnostic d'une complexification des pratiques informationnelles aujourd'hui, l'expérimentation méthodologique et, enfin, l'existence de problèmes épistémologiques.

La première dimension a trait à l'opacité grandissante des phénomènes empiriques en lien avec l'information d'actualité en ligne. Plusieurs paramètres concourent à rendre difficilement intelligibles ces dernières, en l'occurrence, la diversification des sources d'information, la fragmentation de la consommation médiatique, la multiplication des supports matériels, la généralisation du *multitasking* et l'instabilité des pratiques. Ce contexte incite alors – et c'est la deuxième dimension ressortant avec évidence de nos entretiens – les chercheurs à s'engager dans « un bricolage méthodologique », cette expression renvoyant non pas à des errements approximatifs, mais à des pratiques faisant appel à l'ingéniosité et à l'inventivité. La complexification des manières de s'informer invite en effet non seulement à réévaluer rigoureusement les techniques usuelles de collecte et d'analyse de données en sciences sociales, mais aussi à élaborer de nouvelles stratégies méthodologiques. Les experts de notre panel ont, à ce propos, plaidé avec insistance en faveur d'une combinaison/hybridation des approches qualitatives et computationnelles pour appréhender les pratiques informationnelles contemporaines. La dernière dimension relève de considérations épistémologiques. Au regard de ce que d'aucuns appellent le « tournant computationnel » en sciences sociales, les chercheurs de notre panel ont fréquemment opéré une distinction entre les stratégies méthodologiques permettant de mesurer les manières de s'informer actuelles et celles qui autorisent une compréhension de ces dernières. Déterminante, cette dichotomie offre la possibilité de prendre conscience de l'existence de plusieurs modes d'intelligibilité dans les travaux en sciences sociales, d'une part, et de la nécessité de ne pas



substituer un peu trop hâtivement la mesure à la compréhension, d'autre part. L'existence de ce « tournant computationnel » en sciences sociales a elle aussi été mise en question dans nos entretiens. Cette interrogation a fait apparaître des positions pour le moins contrastées parmi les chercheurs. Enfin, certaines personnes que nous avons interrogées ont évoqué les problèmes posés par une discussion a-disciplinaire sur les méthodes de recherche en lien avec le numérique. L'enjeu était alors de mettre en question la légitimité d'une autonomisation de ces débats par rapport aux grandes problématiques propres aux différentes disciplines en sciences sociales.

À partir d'une revue de la littérature savante, nous construisons dans le chapitre III une typologie des principales méthodologies dans les recherches sur les pratiques informationnelles actuelles. Pour ce faire, nous nous sommes inspirés d'une distinction développée par Noortje Marres (2012) dans ses travaux appartenant au domaine des études sur la science et la technologie (*Science and Technology Studies*). Quatre grandes approches méthodologiques ont ainsi été distinguées : les méthodes conventionnelles, l'ethnographie en ligne, les méthodes computationnelles appliquées aux traces des *big data* et les méthodes numériques (*digital methods*). Si les méthodes conventionnelles – qualitatives et quantitatives – sont largement diffusées dans les formations universitaires en sciences sociales ou dans des manuels et, à ce titre, bien connues, il n'en est pas de même pour les trois autres idéaux-types de méthode en lien avec le numérique. L'ethnographie en ligne, que Christine Hine a beaucoup contribué à populariser, vise à produire des « descriptions denses et détaillées » de la réalité observée en prenant acte de la prégnance de la médiatisation numérique dans les pratiques sociales contemporaines. L'ethnographie en ligne s'inscrit dans la continuité de l'ethnographie traditionnelle, mais elle s'efforce d'adapter ses techniques d'enquête au contexte actuel. Cette approche méthodologique se caractérise par ailleurs par sa pluralité, son adaptabilité et sa nature qualitative. Les méthodes computationnelles exploitant les *big data* – les *big methods* pour reprendre l'expression de Noortje Marres – sont nées de la rencontre des puissantes technologies computationnelles et de la disponibilité des traces des *big data*. Ces dernières se caractérisent par leur volume, leur vélocité, leur exhaustivité, la finesse de leur granularité, ainsi que leur flexibi-

lité. Ce troisième idéal-type de méthode s'accompagne par ailleurs de l'utilisation d'outils de visualisation pour rendre compte des traces extraites des plateformes en ligne. L'analyse des traces des *big data*, dans le cadre de ces méthodes computationnelles, repose enfin principalement sur la recherche de corrélations. Le quatrième idéal-type renvoie aux méthodes numériques. Quoique variées, ces dernières ont pour dénominateur commun de vouloir se saisir des opportunités offertes par les objets nativement numériques, par le désir d'articuler des approches qualitatives et quantitatives, mais également par l'ambition de servir la réalisation de recherches en sciences sociales. Nous avons distingué deux grandes familles de méthodes numériques, à savoir les méthodes numériques telles qu'elles sont définies par la *Digital Methods Initiative* coordonnée par le professeur Richard Rogers de l'Université d'Amsterdam, d'une part, et les méthodes numériques quali-quantitatives, d'autre part.

Pour chacun de ces idéaux-types, nous avons repéré une étude de cas concernant les manières de s'informer dans le monde contemporain marqué par le numérique ; nous avons procédé à une présentation analytique de chacune de ces études, ce qui a constitué la matière du chapitre IV. Il s'agissait d'examiner, à partir de recherches d'ores et déjà publiées sur les pratiques informationnelles actuelles, les modalités concrètes de ces différentes stratégies méthodologiques en mettant plus spécifiquement l'accent sur leurs potentialités et limites respectives. Nous avons opté pour une grille d'analyse comportant cinq dimensions, à savoir les présupposés épistémologiques de ces recherches, leurs assises théoriques, leurs techniques de collecte et d'analyse, un aperçu des résultats obtenus et les limites du dispositif technologique. Pour illustrer le recours aux méthodes conventionnelles, nous avons choisi une étude du Pew Research Center sur les usages des médias socionumériques (méthodes quantitatives) et une recherche menée par Guillaume Le Saulnier sur la lecture de la presse en ligne (méthodes qualitatives). par Coralie Le Caroff, dans le cadre de sa thèse de doctorat, ont été mobilisées pour exemplifier l'ethnographie en ligne. Ce travail portait sur les usages sociopolitiques de l'actualité par la médiation de Facebook. Pour mettre en lumière le recours aux méthodes computationnelles appliquées aux *big data* – les *big methods* – nous avons choisi une recherche réalisée en Australie par Axel Bruns, Tim Highfield

et Stephen Harrington. Ce travail concernait le partage de l'information d'actualité par la médiation de Twitter. Dans le cas des méthodes numériques, nous nous sommes focalisés sur les méthodes numériques quali-quantitatives en nous intéressant à une recherche réalisée par Irène Bastard sur le partage de l'information d'actualité par Facebook.

## **Une analyse comparative des idéaux-types de méthode**

Pour procéder à une comparaison entre nos idéaux-types de méthode, nous proposons à nos lecteurs un tableau récapitulant de manière synthétique leurs différentes propriétés. Six dimensions sont considérées dans ce dernier : les ressources financières requises ; les compétences exigées pour mettre en œuvre la méthode ; le type de données principalement manipulées ; le degré de dépendance dans les modalités d'accès aux données de la recherche (l'accès aux données par les chercheurs est-il subordonné aux décisions d'une organisation tierce ?) ; la possibilité de généraliser les résultats de la recherche au regard de son dispositif méthodologique ; et le régime d'intelligibilité propre à la méthode (d'après les distinctions classiques de l'épistémologie).

Tableau de synthèse

	Méthodes conventionnelles quantitatives	Méthodes conventionnelles qualitatives	Ethnographie en ligne	Méthodes computationnelles appliquées aux <i>big data</i>	Méthodes numériques quali-quantitatives
<b>Ressources financières</b>	Élevées	Faibles	Faibles	Élevées	Élevées
<b>Compétences requises</b>	Sciences sociales (statistiques)	Sciences sociales (techniques d'enquête de terrain)	Sciences sociales (techniques d'enquête de terrain)	Ingénierie et recherche informatique	Sciences sociales (techniques d'enquête de terrain) et ingénierie informatique
<b>Types de données</b>	Données collectées par l'administration de questionnaires hors ligne	Données collectées par des techniques d'enquête de terrain hors ligne	Données collectées par des techniques d'enquête de terrain en ligne (et de manière complémentaire hors ligne)	Traces	Données collectées par l'administration de questionnaires et de techniques d'enquête de terrain hors ligne ; traces.
<b>Degré de dépendance des chercheurs pour l'accès aux données</b>	Autonomie des chercheurs	Autonomie des chercheurs	Autonomie des chercheurs	Hétéronomie (dépendance par rapport à la politique des propriétaires de plateforme en ligne)	Hétéronomie (dépendance par rapport à la politique des propriétaires de plateforme en ligne)
<b>Conditions de généralisation des résultats</b>	Avec un échantillon représentatif	En croisant les résultats de la recherche avec d'autres enquêtes de terrain.	En croisant les résultats de la recherche avec d'autres enquêtes de terrain.	Avec un échantillon représentatif par approximation	Avec un échantillon représentatif par approximation et un croisement des résultats de l'enquête avec d'autres enquêtes de terrain
<b>Types d'intelligibilité</b>	Description par la mesure	Compréhension	Compréhension	Description par la mesure	Description par la mesure et compréhension

De ce tableau de synthèse, plusieurs conclusions semblent se dégager. Pour commencer, rien ne nous autorise à affirmer l'existence d'une méthode « ultime », qui jouirait des qualités nécessaires pour répondre à la totalité des problématiques de recherche sur les pratiques informationnelles actuelles. Chaque méthode possède, en réalité, un champ d'application légitime. Par ailleurs, lorsqu'ils mobilisent des techniques de collecte et d'analyse des données, les scientifiques doivent impérativement prendre en compte les situations concrètes dans lesquelles ils effectuent leurs travaux. Chaque contexte s'accompagne de contingences plus ou moins (in)compatibles avec les conditions de flexibilité des techniques de collecte et d'analyse des données évoquées ci-dessus. La première leçon à tirer consiste, par conséquent, à se garder de l'illusion selon laquelle il existerait une méthode susceptible d'éclairer simultanément toutes les dimensions de la réalité des manières de s'informer et qui serait appropriée à toutes les circonstances.

Plus précisément, les méthodes conventionnelles quantitatives, computationnelles, et numériques quali-quantitatives apparaissent adéquates pour mesurer des pratiques sociales à partir d'échantillons de (très) grande taille. De ce fait, elles sont propices pour décrire ou pour expliquer des tendances majeures qui caractérisent notre présent, ou encore pour dresser des états des lieux. Ceci n'implique toutefois pas que les échantillons soient représentatifs : un travail de correction des distorsions – entre, par exemple, un échantillon non probabiliste et la population qu'il vise à « représenter » – est nécessaire. Ce travail correctif est réalisé à partir de la prise en compte de pondérations qui feront en sorte que l'échantillon devienne probabiliste et

représente alors une diversité des catégories d'individus s'approchant de celle de la population totale<sup>115</sup>.

Autant la représentativité peut s'avérer accessible avec les méthodes conventionnelles quantitatives, autant elle ne peut être qu'au mieux approximative avec les méthodes computationnelles et numériques quali-quantitatives<sup>116</sup>. Pour leur part, les méthodes conventionnelles qualitatives, l'ethnographie en ligne et les méthodes numériques quali-quantitatives semblent quant à elles plus adaptées à la compréhension du comportement des acteurs et à la description fine des conditions singulières dans lesquelles ceux-ci agissent. Ces différentes stratégies méthodologiques, en raison du type d'intelligibilité qu'elles confèrent aux pratiques informationnelles contemporaines, rencontrent ainsi des limitations par rapport aux problématiques de recherche auxquelles elles peuvent prétendre apporter des réponses.

Par ailleurs, les méthodes conventionnelles quantitatives, computationnelles et numériques quali-quantitatives requièrent non seulement (la plupart du temps) des ressources financières importantes, mais elles obligent aussi à réunir, au sein d'une équipe, des chercheurs en sciences sociales et des informaticiens (idéalement : chercheurs en informatique et ingénieurs en informatique). Ces contraintes peuvent ainsi se révéler insurmontables pour des organisations dotées de faibles moyens financiers ou n'ayant pas la capacité d'enrôler des chercheurs de profils variés. Nous aimerions toutefois nuancer quelque peu cette proposition pour ce qui est des méthodes computationnelles appliquées aux *big data* et des méthodes qui relèvent des approches numériques quali-quantitatives. Plusieurs facteurs sont en effet susceptibles

---

115. « [...] un échantillon est représentatif dès qu'est vérifiée la double condition ci-dessus : probabilités non nulles et connues. Si ces probabilités sont inégales, certaines catégories d'individus sont proportionnellement plus abondantes dans l'échantillon que dans la population : on dit qu'elles sont sur-représentées. Corrélativement, d'autres sont sous-représentées. [...] il est possible [de] compenser [ces] distorsions : il suffit de dilater convenablement les parties les plus réduites. Ceci s'obtient très simplement en multipliant chacune des observations par l'inverse de la probabilité que l'individu correspondant avait de figurer dans l'échantillon (ce qui est possible, puisque cette probabilité est connue). » (Padieu, 1974, p. 66)

116. Les corpus de type *big data* ne sont pas construits en effet en fonction des critères habituels de représentativité de la théorie statistique des probabilités.

d'affecter le coût du recours à ces deux stratégies méthodologiques. Lorsqu'elles requièrent d'acheter des données numériques aux firmes propriétaires des plateformes en ligne, de développer de nouveaux logiciels permettant de procéder à la manipulation des traces ou de payer des prestataires extérieurs (lorsqu'il est impossible, par exemple, de s'assurer de la collaboration entre plusieurs départements universitaires), les méthodes computationnelles et numériques quali-quantitatives exigent des budgets conséquents. Si tel n'est pas le cas, ces approches méthodologiques peuvent en revanche se montrer plus abordables d'un point de vue économique, et donc accessibles à un plus grand nombre d'organisations dédiées à la recherche.

Une autre ligne de partage distingue les stratégies méthodologiques entre elles. Il s'agit de la nature des données manipulées par ces dernières. Au-delà de la dichotomie entre données hors ligne et données en ligne, les méthodes computationnelles exploitant les *big data* et les méthodes numériques quali-quantitatives ont la particularité de collecter et d'analyser des corpus volumineux de traces. Or, plusieurs chercheurs voient dans celles-ci un moyen d'accéder à « la réalité effective » des pratiques informationnelles et de dépasser, par là même, les erreurs éventuelles des données déclaratives (recueillies à l'aide de questionnaires, d'entretiens, etc.). Tout se passe comme si les traces donnaient la possibilité de se libérer, pour citer une formule ironique de Bourdieu, de la malédiction d'étudier « un objet qui parle » (Bourdieu, Chamboredon et Passeron, 1968, p. 56-57). Les traces seraient, dans cette perspective, ce par quoi les scientifiques pourraient se soustraire aux problèmes de mémoire des enquêtés, à leurs imprécisions ou même à leurs incompréhensions. Les présupposés épistémologiques de cette division entre données *déclaratives* et données *effectives* mériteraient d'être examinés plus longuement, mais il faut reconnaître que cette dernière n'en reste pas moins structurante dans la littérature scientifique.<sup>117</sup>.

---

117. Cette distinction entre les données déclaratives et les traces semble réintroduire, par « la petite porte », la thèse bachelardienne de la rupture épistémologique, chère à la sociologie de Pierre Bourdieu (Bourdieu, Chamboredon et Passeron, 1968, p. 27-49).

Le critère de l'accessibilité des données ne doit pas être négligé pour apprécier ce qui distingue les stratégies méthodologiques entre elles. La collecte et l'analyse des traces, menées dans le contexte des méthodes computationnelles et numériques quali-quantitatives, placent nécessairement les équipes de recherche dans une situation de dépendance par rapport aux organisations propriétaires des plateformes en ligne. Les modalités d'accès aux données de type *big data* sont en effet susceptibles d'évoluer, sans que les chercheurs en soient avertis, conformément aux choix économiques et financiers des propriétaires de ces médias numériques. Cela signifie qu'une étude sur les pratiques informationnelles recourant aux deux approches méthodologiques mentionnées ci-dessus peut être interrompue en raison de changements sur le plan des modalités d'accès, ou elle peut tout simplement s'avérer impossible à réaliser. À l'inverse, les méthodes conventionnelles qualitatives et quantitatives apparaissent préservées de « ces aléas de la fortune », puisqu'elles sont au principe de la construction de leurs propres données. Elles ne mobilisent pas de données d'emprunt.

## Recommandations au Centre d'études sur les médias

À partir de ce qui précède, nous aimerions soumettre six recommandations au Centre d'études sur les médias concernant les dispositifs méthodologiques des recherches en sciences sociales sur les pratiques informationnelles actuelles. Ces recommandations demeurent volontairement générales et ne sauraient être associées à des préconisations à respecter strictement. Dans la mesure où les techniques de collecte et d'analyse de données varient nécessairement en fonction de la problématique et du contexte de la recherche, il ne paraît pas souhaitable d'enfermer la réalisation des études sur les manières de s'informer dans des modèles rigides. Insistons sur ce point : il n'existe pas de méthode de recherche « universelle » permettant de traiter toutes les questions scientifiques ou embrassant toutes les dimensions des pratiques informationnelles.

1. Le développement de stratégies méthodologiques hybrides, articulant des techniques d'enquête qualitatives et quantitatives, nous apparaît plus que souhaitable pour permettre



- la description, la mesure et la compréhension des pratiques informationnelles actuelles.
2. Les méthodes numériques (*digital methods*) constituent une voie prometteuse pour l'étude des manières de s'informer aujourd'hui. Nous préconisons par conséquent une articulation entre les méthodes conventionnelles, que celles-ci soient qualitatives ou quantitatives, et les méthodes computationnelles appliquées aux traces des *big data*.
  3. Dans la mesure où les méthodes numériques (*digital methods*) ouvrent des perspectives intéressantes pour la recherche sur l'information médiatique et le journalisme, nous suggérons la formation d'équipes interdisciplinaires, rassemblant des chercheurs en sciences sociales, en informatique et en sciences des données (*data scientists*)<sup>118</sup>.
  4. En cohérence avec les recommandations précédentes, nous décourageons le recours unique et exclusif à des méthodes computationnelles appliquées aux traces des *big data*. Les *big methods* ne peuvent qu'apporter aux manières de s'informer contemporaines une intelligibilité limitée, fondée sur la mesure, mais déconnectée des logiques de signification des acteurs ou des contextes spécifiques dans lesquels prennent place leurs comportements.
  5. La mobilisation des méthodes numériques (*digital methods*) n'est pas sans difficulté. Certains risques doivent être pris en considération avant l'adoption d'une telle stratégie méthodologique. La volatilité du design des plateformes en ligne constitue un premier problème, dans la mesure où les chercheurs n'ont d'autre choix que de « subir » les évolutions rapides – et parfois radicales – des caractéristiques sociotechniques des médias numériques. De la même manière, les chercheurs doivent avoir conscience

---

118. Le Centre d'études sur les médias (CEM) pourrait par exemple s'inspirer de deux projets de recherche reposant sur des méthodes numériques exécutées par des équipes interdisciplinaires, en l'occurrence une étude sur le phénomène des « *fake news* » (<https://wiki.digitalmethods.net/Dmi/FakeNewsSprint>) et une étude sur la qualité de l'information médiatique au temps de l'intelligence artificielle (<https://medium.com/berkman-klein-center/understanding-media-and-information-quality-in-an-age-of-artificial-intelligence-automation-a562620c7039>).

des bouleversements que peuvent engendrer des changements sur le plan des modalités d'accès aux données des plateformes en ligne.

6. Sur le plan éthique, nous préconisons la mise en place de dispositifs méthodologiques respectueux de la protection des données personnelles des enquêtés, d'une part, et passant par l'obtention du consentement des individus, d'autre part.

# Glossaire

**Big data.** « Le terme fait référence à la fois au volume de données disponibles pour des calculs de tous types et à leur variété, puisqu'il est désormais possible de combiner des données enregistrées par des services officiels (comme les demandes adressées à Pôle emploi), des traces issues d'activité sur les réseaux (des requêtes via Google pour réaliser un CV) et des verbatim, récupérés sur des forums ou des sites Internet de presse par exemple. Leur mise à jour constante (vélocité) grâce aux réseaux confère à ces données un statut différent de celui des bases de données dont les sciences sociales sont familières. Le terme fait aussi référence aux méthodes de calcul qui utilisent ces sources de données et qui relèvent du *machine learning* et de l'inférence à partir de corrélations testées en masse. » (Boullier, 2015b).

**Big methods.** Cette appellation provient d'un nouveau domaine développé au sein des sciences naturelles (mathématiques, physique, informatique) et désigné comme étant la « nouvelle science des réseaux » (*new network science*). Ce nouveau domaine scientifique prend acte de la prolifération des données produites par les plateformes numériques. En effet, dans la dernière décennie, les industries des moteurs de recherche, des jeux vidéo, des réseaux socionumériques, de même que les fournisseurs de services internet ont généré la production d'immenses bases de « données transactionnelles » (Savage et Burrows, 2007). La présence de ces très grandes bases de données (*big data*), combinée à l'émergence de nouvelles techniques de modélisation et de visualisation des données ainsi que d'analyses textuelles sophistiquées, est propice à l'élaboration de nouvelles approches scientifiques (*big methods*). Ces nouvelles approches empiriques (*new network science*) visant à remplacer les techniques classiques de l'enquête par sondage, tendent à maximiser le rôle des techniques mathématiques quantitatives au détriment d'une prise en compte de données qualitatives fournies « subjectivement » par les sujets humains (Marres, 2012, p. 150). Les tenants des *big methods* adoptent une posture positiviste qui prétend pouvoir faire l'économie des sciences humaines et sociales.

**Digital methods (méthodes numériques).** Les approches empiriques rassemblées sous ce libellé s'ancrent dans une proposition théorique formulée en 2009 par Richard Rogers en vue de réorienter le domaine des recherches liées à internet (Rogers, 2009). Il est question pour le chercheur de capter et de s'appropriier – tout en les

recadrant à la lumière de problématiques sociétales (*social issues*) – les méthodes enchâssées déjà dans les dispositifs en ligne (moteurs de recherche, plateformes de médias sociaux numériques). Exemples : analyse comparative des réponses aux mêmes questions à un même moteur de recherche formulées par des internautes situés dans des lieux géographiques différents ; analyse de résultats statistiques obtenus de Google Analytics ; comptage des hyperliens associés à un thème ou à un site particulier ; comptage des Tweets pour un thème choisi et une période donnée ; comparaison de la même page Wikipédia rédigée dans des pays différents ; analyse sociohistorique d'un site archivé à différents moments historiques ; classement et ordonnancement des sites par les moteurs de recherche. Il ne s'agit pas d'utiliser ces dispositifs numériques pour simplement décrire les usages d'internet caractérisés et comptabilisés par ces outils natifs (usage de premier degré). Il s'agit plutôt d'interpréter les résultats obtenus par ces outils pour ce qu'ils disent de la société et des collectifs qui la constituent (dans ses dimensions sociales, culturelles, économiques et politiques). Un usage de second degré (*repurposing*) en quelque sorte. Cette réorientation par les chercheurs des usages des méthodes enchâssées dans le « medium internet » permet ainsi de produire un nouveau type de diagnostics culturels, politiques et sociétaux.

**Digital/numérique.** « Dans un texte publié le 14 janvier 2014 dans le journal *Le Monde*, Marlène Duretz raconte sa quête de sens à propos du couple digital/numérique. Interrogeant son moteur de recherche, l'encyclopédie Wikipédia l'informe que "bien que les terminologies officielles française et québécoise préfèrent 'numérique', l'usage de 'digital' en français se perpétue". Pour l'adjectif "digital", écrit-elle, le Trésor de la langue française propose deux pistes définitionnelles : "relatif au doigt" (*digitus*) ou "qui est exprimé par un nombre" (*numerus*). De son côté, le Grand Robert définit "digital" comme ce "qui appartient aux doigts" : ici nulle trace d'une synonymie avec le "numérique". En langue française, le terme "digital" semble décidément relié au "doigt"...

« Le même Grand Robert définit l'adjectif "numérique" comme étant "ce qui est représenté par un nombre, qui concerne les nombres arithmétiques". Sans doute sous l'emprise d'un esprit du temps qui s'affirme très souvent "numérique", coup de théâtre en 2013 : le Larousse.fr indique que l'adjectif "digital" est considéré dorénavant comme un "synonyme de numérique". Puis, en 2014, nouveau saut qualitatif : le Petit Larousse illustré (version papier) déconseille le terme "digital" et propose l'usage de l'adjectif "numérique". Marlène Duretz – comme d'ailleurs Thibaut de Jaegher de l'*Usine digitale* qui rappelle que "le numérique se pratique avec les doigts" (clavier, souris, écran tactile) – aurait tendance à penser qu'il s'agit d'un "faux combat" sémantique : l'Académie française n'a-t-elle pas rappelé, dans une note du 7 novembre 2013, de ne pas confondre

“ces deux adjectifs qui appartiennent à des langues différentes et dont les sens ne se recouvrent pas”.

« Ce faux combat sémantique écarte pourtant la question du traitement technique du signal électrique dans les réseaux de transmission, d’abord en téléphonie puis en télécommunication. Dans le prolongement des travaux réalisés dans les années 1940 par l’ingénieur et mathématicien Claude Shannon – un théoricien de l’entropie encore davantage que de l’information – des ingénieurs des Bell Labs mettent au point un mode de traitement automatisé du signal qui consiste à transformer l’impulsion électrique en chiffre (“*digit*”), donc à numériser le signal (en anglais, on dira “*digital*” en opposition à “*analog*”) afin d’accroître la performativité des réseaux techniques de communication. Ici se trouve peut-être la racine de la confusion entourant l’usage en français du terme “digital” : l’adjectif deviendrait un anglicisme lorsqu’il serait considéré comme synonyme de “numérique”. C’est ce processus de digitalisation (numérisation) du signal qui rendra possible, dans les décennies 1960-1970, la formidable convergence entre les technologies du transport (télécommunication) et celles de la computation (informatique). » (Proulx, 2016, p. 267-268)

**Trace numérique.** « Dans la culture numérique, le signe, le message et le document sont appelés à être subsumés dans la catégorie des traces. Celle-ci ne désigne pas un nouveau type d’objet, mais un mode inédit de présence et d’efficacité, lié aux caractéristiques techniques et sociales des réseaux. [...] Le signe, tel que les sciences de l’information et de la communication (SIC) et la sémiologie le définissent, procède d’un acte d’énonciation doté de sens et (en partie au moins) d’intentionnalité. L’empreinte numérique, elle, est automatiquement produite à l’occasion d’un calcul, d’un codage ou d’une connexion, le plus souvent sans que le sujet en soit conscient. Au lieu d’articuler une face sensible (signifiant) à une représentation psychique (signifié), la trace assigne une signature invisible à un comportement informationnel, qui n’est pas toujours perçu comme tel. Téléphoner, voyager, cliquer sur un lien, commander un produit en ligne... autant d’activités que l’on pratique “en aveugle”, sans les éprouver comme traçage. [...] Quand chaque agissement social se traduit en données, il n’est plus besoin de poser le cadre d’une relation interpersonnelle pour établir qu’on ne peut pas communiquer, puisque désormais on ne peut plus ne pas laisser de traces. » (Merzeau, 2009, p. 23-24).



# Bibliographie

- Akrich, M., Callon, M., et Latour, B. (dir.) (2006). *Sociologie de la traduction: Textes fondateurs*. Paris: Presses des Mines.
- Allard, L., et Vandenberghe, F. (2003). Express yourself! Les pages perso. *Réseaux*, 1(117), 191-219.
- Andel, P.V. et Bourcier, D. (dir.) (2009). *La sérendipité dans les sciences, les arts et la décision*. Communication présentée au Colloque de Cerisy, Cerisy-la-salle.
- Anderson, C. (2008). The End of Theory. Will the Data Deluge Makes the Scientific Method Obsolete? *Wired*: <https://www.wired.com/2008/06/pb-theory/> (consulté le 13-01-2018).
- Ansart, p. (1999). Explication. Dans *Dictionnaire de sociologie*. Paris: Le Robert / Seuil, 212-213.
- Auray, N. (2016). *L'Alerte ou l'Enquête. Une sociologie pragmatique du numérique*. Paris: Presses des Mines.
- Bachelard, G. (1995). *Le nouvel esprit scientifique*. Paris: Presses universitaires de France.
- Bastard, I. (2015). *De proches en pages, de pages en proches: Exploration et réception des informations en ligne* (Thèse de doctorat, Paris, Télécom ParisTech).
- Bastard, I., Cardon, D., Charbey, R., Cointet, J. P., et Prieur, C. (2017). Facebook, pour quoi faire?. *Sociologie*, 8(1), 57-82.
- Bastard, I., Cardon, D., Fouetillou, G., Prieur, C. et Roux, S. (2013). Travail et travailleurs de la donnée. *InternetActu.net*, FING: <http://www.internetactu.net/2013/12/13/travail-et-travailleurs-de-la-donnee/> (repéré le 14 avril 2018).
- Bastin, G. et Tubaro, p. (2018). Le moment *big data* des sciences sociales. *Revue française de sociologie*, 59(3), 375-394.
- Beaudouin, V. et Denis, J (2014). *Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique*. Repéré à <https://halshs.archives-ouvertes.fr/halshs-01078530>
- Becker, H. S. (1960). Notes on the Concept of Commitment. *The American Journal of Sociology*, 66 (1), p. 32-40. Traduit en français in *Tracés*, n° 11, 2006 (1), p. 177-192.
- Becker, H. S. (2002). *Les ficelles du métier. Comment conduire sa recherche en sciences sociales*. Paris: Éditions La Découverte.
- Becker, H. S. (2004). Épistémologie de la recherche qualitative. Dans A. Blanc et A. Pessin (dir.), *L'art du terrain. Mélanges offerts à Howard S. Becker* (p. 58-89). Paris: L'Harmattan.

- Béliard A. S. et Le Caroff, C. (2018). L'archivage et l'observation de la participation en ligne à l'épreuve d'une approche ethnographique. Dans S. Lecossais et Quemener N. (dir.), *En quête d'archives. Bricolages méthodologiques en terrains médiatiques* (n.d). Paris: INA Éditions.
- Béréni, L., Chauvin, S., Jaunait, A., et Revillard, A. (2008). *Introduction aux gender studies: manuel des études sur le genre*. Louvain-La-Neuve: De Boeck.
- Beuscart, J. S., Dagiral, É., et Parasio, S. (2016). *Sociologie d'internet*. Paris: Armand Colin.
- Blin, T. (2010). *Requiem pour une phénoménologie. Sur Alfred Schütz, Merleau-Ponty et quelques autres*. Paris: Éditions du Félin.
- Blumler, J.G. et Katz, E. (dir.) (1974). *The Uses of Mass Communications. Current Perspectives on Gratifications Research*. Beverly Hills, CA: Sage Publications.
- Boullier, D. (2004). La fabrique de l'opinion publique dans les conversations télé. *Réseaux*, 4(126), 57-87.
- Boullier, D. (2015a). Vie et mort des sciences sociales avec le Big Data. *Socio*, 4, 19-37.
- Boullier, D. (2015b). Les sciences sociales face aux traces du big data. *Revue française de science politique*, 65(5), 805-828
- Boullier, D. (2016). *Sociologie du numérique*. Paris: Armand Colin.
- Bourdieu, P., Chamboredon, J. C. et Passeron, J. C. (1968). *Le métier de sociologue*. Paris: Mouton.
- Boyd, D., et Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.
- Breton, P., et Proulx, S. (2012). *L'explosion de la communication. Introduction aux théories et aux pratiques de la communication*. Paris: La Découverte.
- Brown, W. (2002). At the edge. *Political Theory*, 30(4), 556-576.
- Bruns, A. (2005). *Gatewatching: Collaborative online news production*. New York: Peter Lang.
- Bruns, A., Burgess, J. E., Crawford, K., et Shaw, F. (2012). #qldfloods and @QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods. Repéré à <http://eprints.qut.edu.au/48241/>
- Bruns, A., et Burgess, J. (2011). #ausvotes: How Twitter covered the 2010 Australian federal election. *Communication, Politics & Culture*, 44(2), 37-56.
- Bruns, A., et Burgess, J. (2012a). Researching news discussion on Twitter: New methodologies. *Journalism Studies*, 13(5-6), 801-814.
- Bruns, A., et Burgess, J. (2012b). *Notes towards the scientific study of public communication on Twitter*. Repéré à [https://eprints.qut.edu.au/55916/1/Notes\\_towards\\_the\\_Scientific\\_Study\\_of\\_Public\\_Communication\\_on\\_Twitter.pdf](https://eprints.qut.edu.au/55916/1/Notes_towards_the_Scientific_Study_of_Public_Communication_on_Twitter.pdf)
- Bruns, A., et Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(4). Repéré à <http://uncommonculture.org/ojs/index.php/fm/article/view/3937>



- Bruns, A., et Stieglitz, S. (2013). Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology*, 16(2), 91-108.
- Bruns, A., Highfield, T., et Harrington, S. (2013). *Sharing the news: Dissemination of links to Australian news sites on Twitter*. Repéré à <http://snurb.info/files/2013/Sharing%20the%20News.pdf>
- Buchanan, E. (2011). Internet Research Ethics: Past, Present, and Future. Dans C. Mia et C. Ess (dir.), *The Handbook of Internet Studies* (p. 83-108). Malden, MA, Wiley-Blackwell.
- Burgess, J., et Bruns, A. (2012). Twitter archives and the challenges of « Big Social Data » for media and communication research. *M/C Journal*, 15(5). Repéré à <http://journal.media-culture.org.au/index.php/mcjournal/article/view/561Driscoll/0>
- Burrows, R., et Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, April-June 2014 (p. 1-6).
- Carnap, R. (1973). *Les fondements philosophiques de la physique*. Paris: Armand Colin.
- Casilli, A., et Tubaro, p. (2012). Social media censorship in times of political unrest-a social simulation experiment with the UK riots. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 115(1), 5-20.
- Chambat, p. (1994). Usages des technologies de l'information et de la communication (TIC) : évolution des problématiques. *Technologies de l'information et société*, 6(3), 249-270.
- Chapoulie, J. M. (2001). *La tradition sociologique de Chicago 1892-1961*. Paris: Éditions du Seuil.
- Chalmers, A. F. (1990). *Qu'est-ce que la science ?*. Paris: Le livre de poche.
- Coutant, A. (2015). Les approches sociotechniques dans la sociologie des usages en SIC. *Revue française des sciences de l'information et de la communication*, (6). Repéré à <https://journals.openedition.org/rfsic/1271>
- Culturomics (s.d). Repéré à <http://www.culturomics.org/Resources/faq#culturomics>
- Datchary, C. (2011), *La dispersion au travail*. Toulouse: Octarès.
- Desrosières, A. et Kott, S. (2005). Quantifier. *Genèses*, (1), 2-3.
- Dilthey, W. (1992), *Introduction aux sciences de l'esprit*. Paris: Cerf.
- Dodge, Y. (2007). Statistique. Dans *Statistique: dictionnaire encyclopédique*. Paris: Springer Science & Business Media, 478-481.
- Dumitrica, D., et Gaden, G. (2008). Knee-high boots and six-pack abs: Autoethnographic reflections on gender and technology in Second Life. *Journal For Virtual Worlds Research*, 1(3), 3-23.
- Eco, U. (1985). *Lector in fabula*. Paris: Grasset.
- Epstein, J. M., et Axtell, R. (1996). Growing artificial societies: social science from the bottom up. Washington: Brookings Institution Press.

- Ford, H. (2014). Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data & Society*, 1 (2), p. 1-3.
- Foucault, M. (2001). *Dits et écrits II, 1976-1988*. Paris: Gallimard.
- Gallant, N., Latzko-Toth, G., et Pastinelli, M. (2015). *Circulation de l'information sur les médias sociaux pendant la grève étudiante de 2012 au Québec*. Québec: Centre d'études sur les médias. Repéré à <http://espace.inrs.ca/2706/1/CirculationInformation.pdf>
- Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.
- Geertz, C. (2008). Thick description: Toward an interpretive theory of culture. Dans T. Oakes et P. Price (dir.), *The Cultural Geography Reader* (p. 41-51). New York: Routledge.
- Giroux, D., Proulx, S. et Souissi, S. (2017). *Les transformations dans les manières de s'informer au Québec (2011-2016)*. Québec: Centre d'études sur les médias. Repéré à: <http://www.cem.ulaval.ca/pdf/Transformations.pdf>
- Gonzales, p. (1996). Production journalistique et contrat de lecture: autour d'un entretien avec Eliseo Veron. *Quaderni*, 29(1), 51-59.
- Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological theory*, 1, 201-233.
- Habermas, J. (1992). *L'espace public*. Paris: Payot.
- Hammersley, M. (1992). *What's wrong with ETHNOGRAPHY?*. New York: Routledge.
- Héran, F. (1984). L'assise statistique de la sociologie. *Économie et statistique*, n° 168, 23-35.
- Hine, C. (2000). *Virtual Ethnography*. London: Sage.
- Hine, C. (2005). *Virtual Methods*. Oxford: Berg.
- Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Bloomsbury.
- Hine, C. (2017). Ethnographies of Online Communities: Modes, Varieties, Affordances. Dans N. G. Fielding, R. M. Lee et G. Blank (dir.), *The Sage Handbook of Online Research Methods* (2<sup>e</sup> éd., p. 401-415). London: Sage.
- Hirschhorn, M. (1999). Sociologie de la compréhension et méthode compréhensive. Dans *Dictionnaire de sociologie*. Paris: Le Robert / Seuil, 98-99.
- Horkheimer, M. (1974). *Théorie traditionnelle et théorie critique*. Paris: Gallimard.
- Horkheimer, M. (2009). *Théorie critique*. Paris: Payot.
- Jauréguiberry, F., et Proulx, S. (2011). *Usages et enjeux des technologies de communication*. Toulouse: Érès.
- Jouët, J. (1993). Pratiques de communication et figures de la médiation. *Réseaux*, 11(60), 99-120.
- Jouët, J. et Le Caroff, C. (2016). L'observation ethnographique en ligne. Dans C. Barats (dir.), *Manuel d'analyse du web en sciences humaines et sociales* (p. 156-175). Paris: Armand Colin.

- Katz, E. et P. Lazarsfeld (2006). *Influence personnelle. Ce que les gens font des médias*. Paris: Armand Colin.
- Kaufman, J.C. (1996). *L'entretien compréhensif*. Paris: Armand-Colin.
- Kitchin, R. (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12.
- Kotras, B. (2013). Des opinions qui comptent. Influence, visibilité et hiérarchisation des locuteurs sur le web. *E-réputation et traces numériques : dimensions instrumentales et enjeux de société*. Colloque international, Université Toulouse 1 Capitole, 21-23 mars.
- Kroker, A. et M. Kroker (2008). *Critical Digital Studies : A Reader*. Toronto: University of Toronto Press.
- Krüger, N., Süeglit, S., et Potthoff, T. (2012). *Brand Communication In Twitter-A Case Study On Adidas*. Repéré à <https://pdfs.semanticscholar.org/0a5b/c525d4f42f5105a57b6a0a22413c6595c210.pdf>
- Lalande, A. (1993). Politique. Dans *Vocabulaire technique et critique de la philosophie* (Vol. 2, p. 785-786). Paris: Presses universitaires de France.
- Latour, B. (2001). *L'espoir de Pandore : pour une version réaliste de l'activité scientifique*. Paris: La Découverte.
- Latzko-Toth, G. et Proulx, S. (2013). Enjeux éthiques de la recherche sur le Web. Dans C. Barats (dir.), *Manuel d'analyse du web en Sciences Humaines et Sociales* (p. 38-59). Paris: Armand Colin.
- Latzko-Toth, G., Bonneau, C., et Millette, M. (2017). Small data, thick data: Thickening strategies for trace-based social media research. Dans L. Sloan et A. Quan-Hasse (dir.), *The SAGE handbook of social media research methods* (p. 199-214). London: Sage.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... et Jebara, T. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915), 721-723.
- Le Caroff, C. (2015a). *Les usages sociopolitiques de l'actualité en ligne. S'informer, partager et commenter sur Facebook* (Thèse de doctorat, Paris, Université Paris 2).
- Le Caroff, C. (2015b). Le genre et la prise de parole politique sur Facebook. *Participations*, 2(12), 109-137.
- Le Caroff, C. (2018). Le partage de l'actualité politique sur les profils personnels de Facebook. Dans A. Mercier et N. Pignard-Cheyne (dir.), *#info: Commenter et partager l'actualité sur Twitter et Facebook* (p. 199-226). Paris: éditions de la maison des sciences de l'homme.
- Le Saulnier, G. (2013). La lecture de la presse en ligne. L'appropriation des contenus d'actualité au défi de la technique. Dans J. Jouët et R. Rieffel (dir.), *S'informer à l'ère numérique* (p. 57-83). Rennes: Presses universitaires de Rennes.

- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., et Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook. *com. Social networks*, 30(4), 330-342.
- Mallein, P., et Toussaint, Y. (1994). L'intégration sociale des technologies d'information et de communication: une sociologie des usages. *Technologies de l'information et société*, 6(4), 315-335.
- Manovich, L. (2007). *Cultural analytics: Analysis and visualization of large cultural data sets*. Repéré à <http://manovich.net/index.php/projects/cultural-analytics-visualizing-cultural-patterns>.
- Manovich, L. (2011). Trending: The Promises and the Challenges of Big Social Data. Repéré à <http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data>
- Manovich, L. (2015). The Science of Culture? Social Computing, Digital Humanities, and Cultural Analytics. Repéré à <http://manovich.net/index.php/projects/cultural-analytics-social-computing>
- Markham, A. N. (1998). *Life online: Researching real experience in virtual space*. Lanham: Altamira.
- Marres, N. (2012). The redistribution of methods: on intervention in digital social research, broadly conceived. *The Sociological Review*, 60(1), 139-165.
- Marres, N. et E. Weltevrede (2013). Scraping the Social?. *Journal of Cultural Economy*, 6 (3), 313-335.
- Martin, O. (1999). Statistiques sociologiques. Dans *Dictionnaire de sociologie*, Paris: Le Robert / Seuil, 501-504.
- Mellet, K. (2009). Aux sources du marketing viral. *Réseaux*, 5(157-158), 267-292.
- Merzeau, L. (2009). Du signe à la trace: l'information sur mesure. *Hermès*, (53), 21-29.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Pinker, S. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176-182.
- Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, 79(1), 119-157. Repéré à <https://crypto.stanford.edu/portia/papers/RevnissenbaumDTP31.pdf>
- Olivier de Sardan, J. P. (2008). *La rigueur du qualitatif*. Louvain-la-neuve: Bruylant Academia.
- Padieu, R. (1974). Qu'est-ce que la représentativité? *Économie et statistique*, n° 56, 65-66.
- Paillé, p. (1996). De l'analyse qualitative en général et de l'analyse thématique en particulier. *Recherches qualitatives*, (15), 179-194.
- Pew Research Center (2018a). U.S. Survey Research. Repéré à <http://www.pewresearch.org/methodology/u-s-survey-research/>
- Pew Research Center (2018b). Our mission. Repéré à <http://www.pewresearch.org/about/our-mission/>


- Pinch, T. J., et Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, 14(3), 399-441.
- Popper, K. (1973). *La logique de la découverte scientifique*. Paris: Payot.
- Proulx, S. (1994). Les différentes problématiques de l'usage et de l'utilisateur. Dans A. Vitalis (dir.), *Médias et nouvelles technologies. Pour une socio-politique des usages* (p. 127-157). Rennes: Apogée.
- Proulx, S. (2016). Le numérique se pratique avec les doigts. Dans *Le texte à venir*, Classiques Garnier, *Études Digitales*, 1(1), 267-268.
- Proulx, S. (2018). Controverses autour de l'économie dite collaborative : innovation participative, emprise du marché, faiblesse de la régulation. Dans B. Laurent, M.J. Baker, V. Beaudouin et N. Raulet-Croset (dir.), *Innovation et Participation : Approches Critiques* (p. 231-251). Paris: Presses des Mines.
- Proulx, S. et D. Maillet (1998). La construction ethnographique des publics de télévision. Dans S. Proulx (dir.), *Accusé de réception. Le téléspectateur construit par les sciences sociales* (p. 121-161). Québec: Les Presses de l'Université Laval.
- Reuters (2017). *Digital News Report*. Oxford: Oxford University.
- Rey, A. (2012). *Dictionnaire historique de la langue française*. Paris: Le Robert.
- Rogers, R. (2009). The End of the Virtual: Digital Methods. Repéré à [http://www.govcom.org/rogers\\_paris\\_medialab.pdf](http://www.govcom.org/rogers_paris_medialab.pdf).
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA.: The MIT Press.
- Rogers, R. (2015). « Digital methods for web research ». Dans J. Tainter, T. Taylor, R. Brain et J. Lobo (dir.), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*. Toronto: Wiley-BlackWell
- Savage, M., et Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology* 41(5), p. 885-899.
- Smith, A., et Anderson, A. (2018). Social Media Use in 2018. Repéré à [http://assets.pewresearch.org/wp-content/uploads/sites/14/2018/03/01105133/PI\\_2018.03.01\\_Social-Media\\_FINAL.pdf](http://assets.pewresearch.org/wp-content/uploads/sites/14/2018/03/01105133/PI_2018.03.01_Social-Media_FINAL.pdf)
- Soler, L. (2009). *Introduction à l'épistémologie*. Paris: Ellipses.
- Tedjamulia, S. J., Dean, D. L., Olsen, D. R., et Albrecht, C. C. (2005, January). Motivating content contributions to online communities: Toward a more comprehensive theory. Repéré à <https://ieeexplore.ieee.org/abstract/document/1385630/>
- Thomann, X. (2017). L'info vidéo active ses réseaux. *Télérama*, n° 3543, Paris, 6 décembre, p. 30-34.
- Thompson, J. B. (1995). *The Media and Modernity*. Stanford: Stanford University Press.
- Varenne, F. (2007). *Du modèle à la simulation informatique*. Paris: Vrin.

- Varenne, F. (2010). Les simulations computationnelles dans les sciences sociales. *Nouvelles perspectives en Sciences Sociales: revue internationale de systémique complexe et d'études relationnelles*, 5(2), 17-49.
- Varenne, F. (2016). Histoire de la modélisation : quelques jalons. Repéré à : [https://www.researchgate.net/publication/315613805\\_Histoire\\_de\\_la\\_modelisation\\_quelques\\_jalons](https://www.researchgate.net/publication/315613805_Histoire_de_la_modelisation_quelques_jalons)
- Venturini, T. (2012). Great expectations. Repéré à <https://hal.archives-ouvertes.fr/hal-01064258>
- Venturini, T., Cardon, D., et Cointet, J. P. (2014). Présentation. *Réseaux*, 6(188), 9-21.
- Voirol, O. (2011). L'intersubjectivation technique : de l'usage à l'adresse. Pour une théorie critique de la culture numérique. Dans J. Denouël et F. Granjon (dir.), *Communiquer à l'ère numérique. Regards croisés sur la sociologie des usages* (p. 127-157). Paris: Presses des Mines.
- Weber, M. (1965). *Essais sur la théorie de la science*. Repéré à [http://classiques.uqac.ca/classiques/Weber/essais\\_theorie\\_science/essais\\_theorie\\_science.html](http://classiques.uqac.ca/classiques/Weber/essais_theorie_science/essais_theorie_science.html)
- Weltevrede, E. (2016). *Repurposing digital methods: The research affordances of platforms and engines*. Repéré à [https://pure.uva.nl/ws/files/2569586/168511\\_NIET\\_IDENTIEK\\_Weltevrede\\_RepurposingDigitalMethods.pdf](https://pure.uva.nl/ws/files/2569586/168511_NIET_IDENTIEK_Weltevrede_RepurposingDigitalMethods.pdf)
- White, D. M. (2018). « The 'gate keeper': A case study in the selection of news ». In *The Media, Journalism and Democracy* (p. 119-126). New York: Routledge.
- Wouters, C. (2007). *Informalization: Manners and emotions since 1890*. London: Sage.

# Annexe

## Liste des personnes interviewées

- Claire Balleys, Université de Genève, Genève
- Irène Bastard, Bibliothèque nationale de France, Paris
- Valérie Beaudoin, Télécom ParisTech, Paris
- Jean-Samuel Beuscart, Orange Labs, Paris
- Claudine Bonneau, Université du Québec à Montréal
- Dominique Boullier, Polytechnique, Lausanne
- Dominique Cardon, SciencePo, Paris
- Antonio Casilli, EHESS et Télécom ParisTech, Paris
- Raphaël Charbey, Télécom ParisTech, Paris
- Alexandre Coutant, Université du Québec à Montréal
- Éric Dagiral, Université Paris Descartes, Paris
- Jérôme Denis, CSI, Mines ParisTech, Paris
- Célya Gruson-Daniel, chercheure indépendante, Paris
- Sophie Jehel, Université Paris 8, Paris
- Josiane Jouët, Université Paris II-Assas, Paris
- Guillaume Latzko-Toth, Université Laval, Québec
- Jacques-François Marchandise, FING, Paris
- Cécile Méadel, Université Paris II-Assas, Paris
- Florence Millerand, Université du Québec à Montréal
- Mélanie Millette, Université du Québec à Montréal
- Sylvain Parasie, Université de Marne-la-Vallée
- Dominique Pasquier, CNRS et Télécom ParisTech, Paris
- Marc-André Pilon, entrepreneur, Montréal
- Christophe Prieur, Télécom ParisTech, Paris
- Jean-Hugues Roy, Université du Québec à Montréal
- Christine Thoër, Université du Québec à Montréal



Le Centre d'études sur les médias, un organisme sans but lucratif fondé en 1992, est un lieu de recherche, mais il est également un agent de concertation entre les entreprises de communication, les milieux gouvernementaux et universitaires. Il compte trois partenaires universitaires: le Département d'information et de communication de l'Université Laval, l'École des médias de l'UQAM et l'Université de Montréal.

